This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2000, 18:53:45 ; Search time 4457.97 Seconds
(without alignments)
3382.078 Million cell updates/sec

Title: US-09-332-522B-1
Perfect score: 3419
Sequence: 1 ggtttaattacccaagtttg.....taaatacgattatatataaa 3419

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: em_est35:*
83: em_est36:*
84: em_est37:*
85: gb_est48:*
86: gb_est49:*
87: gb_est50:*
88: gb_est51:*
89: gb_est52:*
90: gb_est53:*
91: gb_est54:*
92: gb_est55:*
93: gb_gss1:*
94: gb_gss2:*
95: gb_gss3:*
96: gb_gss4:*
97: em_gss1:*
98: em_gss2:*
99: em_gss3:*
100: em_gss4:*
101: gb_gss5:*
102: gb_gss6:*
103: gb_gss7:*
104: gb_gss8:*
105: gb_gss9:*
106: em_gss5:*
107: em_gss6:*
108: em_gss7:*
109: em_gss8:*
110: em_gss9:*
111: em_gss10:*
112: em_gss11:*
113: gb_gss10:*
114: gb_gss11:*
115: em_gss12:*
116: gb_gss12:*

117: gb_gss13:*
118: gb_gss14:*
119: gb_gss15:*
120: gb_gss16:*
121: gb_gss17:*
122: gb_gss18:*
123: gb_gss19:*
124: em_gss13:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	373.6	10.9	402	81	D32385	D32385 CELK012HZR
	2	359	10.5	360	81	D35004	D35004 CELK015HZF
	3	355	10.4	360	81	D68853	D68853 CELK056DZF
	4	352.8	10.3	378	80	C11998	C11998 C11998 Yuji
	5	350.6	10.3	360	81	D67853	D67853 CELK087G4F
	6	347.4	10.2	360	81	D34815	D34815 CELK012HZF
	7	340	9.9	360	81	C63905	C63905 C63905 Yuji
c	8	326	9.5	329	81	D64806	D64806 CELK012GZR
c	9	305.8	8.9	328	81	D64735	D64735 CELK087G4R
c	10	229.2	6.7	252	80	C10363	C10363 C10363 Yuji
	11	62.8	1.8	337	64	AW115144	AW115144 rs64a09.y
	12	58.4	1.7	506	80	AW701427	AW701427 uq86a04.y
c	13	55.4	1.6	922	122	CNS0073W	AL066784 Drosophil
	14	52.8	1.5	1101	122	CNS00LO0	AL068607 Drosophil
c	15	50.4	1.5	1101	122	CNS0100X	AL098379 Drosophil
	16	49.8	1.5	352	47	AL043416	AL043416 DKFZp434I
	17	49.4	1.4	598	28	AA567826	AA567826 HL01847.5
c	18	49.4	1.4	997	122	CNS005TE	AL060767 Drosophil
c	19	49.4	1.4	1101	122	CNS00LT2	AL078714 Drosophil
	20	48.6	1.4	939	122	CNS00CNG	AL059400 Drosophil
	21	48.4	1.4	1101	122	CNS0039G	AL063921 Drosophil
c	22	48.2	1.4	884	122	CNS006U0	AL065923 Drosophil
c	23	47.8	1.4	895	122	CNS0071A	AL066286 Drosophil
	24	47.6	1.4	554	47	AL040138	AL040138 DKFZp434D
c	25	46.2	1.4	1225	123	CNS0166K	AL106358 Drosophil
	26	45.2	1.3	937	122	CNS006ST	AL065880 Drosophil
c	27	42.6	1.2	987	122	CNS00418	AL066537 Drosophil
	28	42.6	1.2	1101	122	CNS0106X	AL098595 Drosophil
c	29	42.4	1.2	872	123	CNS017WC	AL108582 Drosophil
c	30	42.4	1.2	909	122	CNS00JTL	AL076720 Drosophil
c	31	42	1.2	637	94	AQ159947	AQ159947 mgxb0002H
c	32	41.8	1.2	810	120	B20620	B20620 T24C2-Sp6.1
	33	41.8	1.2	848	122	CNS00GWQ	AL072642 Drosophil
	34	41.8	1.2	1101	122	CNS00ER8	AL069549 Drosophil
c	35	41.6	1.2	977	123	CNS0164E	AL106280 Drosophil
c	36	41.4	1.2	1080	122	CNS00EPP	AL069494 Drosophil
	37	41.4	1.2	1101	123	CNS0181N	AL108773 Drosophil
	38	41.4	1.2	1155	122	CNS007M5	AL067239 Drosophil
c	39	41.2	1.2	902	122	CNS006QP	AL065804 Drosophil
c	40	41	1.2	928	122	CNS0102F	AL098433 Drosophil
c	41	40.6	1.2	443	31	AA806181	AA806181 oc25e08.s
c	42	40.6	1.2	468	114	AQ812374	AQ812374 HS_5267_B
c	43	40.6	1.2	1100	122	CNS00ZBG	AL097462 Drosophil
	44	40.6	1.2	1101	123	CNS0182P	AL108811 Drosophil
c	45	40.4	1.2	1101	122	CNS0039G	AL063921 Drosophil

ALIGNMENTS

RESULT 1
D32385/c
LOCUS D32385 402 bp mRNA EST 08-AUG-1994
DEFINITION CELK012HZR Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA clone yk12h12 3', mRNA sequence.
ACCESSION D32385
VERSION D32385.1 GI:523348
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 402)
AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and Tabara,H.
TITLE Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
FEATURES Location/Qualifiers
source 1. .402
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk12h12"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT 100 a 94 c 98 g 110 t
ORIGIN

Query Match 10.9%; Score 373.6; DB 81; Length 402;
Best Local Similarity 98.2%; Pred. No. 2.1e-99;
Matches 387; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

```
Qy   3002 ggcctacgaatggatta-tgaactcgctgctcgacgcgtggcgttccaatctattcgcat 3060
          | ||||||  |||||||| ||||||| ||||||||||||||||||||||||||||||||
Db    394 GCCCTACGGATGGATTATTGAACTCCCTGCTCGACGCGTGGCGTTCCAATCTATTCGCAT 335

Qy   3061 cgaaaccctactggacacaaagcttcaagggacaatccacgtttagtacgctttatcaag 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    334 CGAAACCCTACTGGACACAAAGCTTCAAGGGACAATCCACGTTTAGTACGCTTTATCAAG 275

Qy   3121 aggcgtataatcattatgcgattattaatgggacaaggggagattgttggagactatttg 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    274 AGGCGTATAATCATTATGCGATTATTAATGGGACAAGGGGAGATTGTTGGAGACTATTTG 215

Qy   3181 tctacgagctcacgtgccgaatgctcaacggagccaacccacaagccacgtggtcaggcg 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    214 TCTACGAGCTCACGTGCCGAATGCTCAACGGAGCCAACCCACAAGCCACGTGGTCAGGCG 155

Qy   3241 yccgacgcgttcgatctacaaaaatggacgcggtccgaggaagagtgagcatgcgacgct 3300
          :|||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    154 TCCGACGCGTTCGATCTACACAAATGGACGCGGTCCGAGGAAGAGTGAGCATGCGACGCT 95

Qy   3301 cggctcaaccggacgcatttcatcttcatacactggttaaactacatacttctatggatc 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     94 CGGCTCAACCGGACGCATTTCATCTTCATACACTGGTTAAACTACATACTTCTATGGATC 35

Qy   3361 tttgaattgaacaaaaaatgattttattcagaat 3394
          |||||||||||||| ||||||||||||||||||
Db     34 TTTGAATTGAACAACAAATGATTTTATTCAGAAT 1
```

RESULT 2
D35004
LOCUS D35004 360 bp mRNA EST 08-AUG-1994
DEFINITION CELK015HZF Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA clone yk15h12 5', mRNA sequence.
ACCESSION D35004
VERSION D35004.1 GI:526520

```
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
            Tabara,H.
  TITLE     Toward an expression map of the C.elegans genome
  JOURNAL   Unpublished (1994)
COMMENT     On Apr 7, 1998 this sequence version replaced gi:3036136.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /db_xref="taxon:6239"
                     /clone="yk15h12"
                     /clone_lib="Yuji Kohara unpublished cDNA"
                     /sex="hermaphrodite, male"
                     /tissue_type="whole animal"
                     /dev_stage="varied"
BASE COUNT      123 a     63 c    100 g     73 t      1 others
ORIGIN
```

```
  Query Match             10.5%;  Score 359;  DB 81;  Length 360;
  Best Local Similarity   99.7%;  Pred. No. 4.2e-95;
  Matches  359;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     773 atcgattgaagatgctccagagacgaaaccaactcatttggttgaaccacaaagtccaaa 832
           |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db       1 ATCGATTGAAGATGCTCCAGAGACGANACCAACTCATTTGGTTGAACCACAAAGTCCAAA 60

Qy     833 aagcccgcagaatatgaaagaggagcttcttcggttactagttaacatgtctccgagtga 892
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AAGCCCGCAGAATATGAAAGAGGAGCTTCTTCGGTTACTAGTTAACATGTCTCCGAGTGA 120

Qy     893 agttgaacggttaaagaataaaaaatcaggagcatgttcagcgacgaatgggccatcgag 952
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AGTTGAACGGTTAAAGAATAAAAAATCAGGAGCATGTTCAGCGACGAATGGGCCATCGAG 180

Qy     953 gagtaaggagaaggcggcgaagattgtgattcaggagacagcggaaggggatgaagatga 1012
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GAGTAAGGAGAAGGCGGCGAAGATTGTGATTCAGGAGACAGCGGAAGGGGATGAAGATGA 240

Qy    1013 ggatgatgaggatagtgattccggggagactatgtctcagggaactactattattgttcg 1072
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GGATGATGAGGATAGTGATTCCGGGGAGACTATGTCTCAGGGAACTACTATTATTGTTCG 300

Qy    1073 aagaccaaaaaccgagcgtcgtacggcacacaatctcatcgaaaagaagtatagatgctc 1132
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AAGACCAAAAACCGAGCGTCGTACGGCACACAATCTCATCGAAAAGAAGTATAGATGCTC 360
```

```
RESULT   3
D68853
LOCUS       D68853         360 bp    mRNA            EST       07-DEC-1995
DEFINITION  CELK056DZF Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk56d12 5', mRNA sequence.
ACCESSION   D68853
VERSION     D68853.1  GI:1104494
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
            Tabara,H.
  TITLE     Toward an expression map of the C.elegans genome
  JOURNAL   Unpublished (1994)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3188359.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /db_xref="taxon:6239"
                     /clone="yk56d12"
                     /clone_lib="Yuji Kohara unpublished cDNA"
                     /sex="hermaphrodite, male"
                     /tissue_type="whole animal"
                     /dev_stage="varied"
BASE COUNT       97 a     72 c     96 g     90 t      5 others
ORIGIN
```

```
  Query Match             10.4%;  Score 355;  DB 81;  Length 360;
  Best Local Similarity   98.6%;  Pred. No. 6.4e-94;
  Matches  355;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy    2540 cctaaatgacaatgaccgtctcgatgatgaagtagacgtggttgatgtttcaagactttt 2599
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CCTAAATGACAATGACCGTCTCGATGATGAAGTAGACGTGGTTGATGTTTCAAGACTTTT 60

Qy    2600 ggtgacaatttcaacgcagtgcgctgccattttgactaatgagaaggatgagtcagcgaa 2659
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GGTGACAATTTCAACGCAGTGCGCTGCCATTTTGACTAATGAGAAGGATGAGTCAGCGAA 120

Qy    2660 attcggaacctggatctctcgaaacggagatgcttgttgcacatggtggacgcacgttct 2719
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ATTCGGAACCTGGATCTCTCGAAACGGAGATGCTTGTTGCACATGGTGGACGCACGTTCT 180

Qy    2720 gacatgtggaatctattggaggagtaacaagaatgagctggcacggcaacactattcact 2779
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GACATGTGGAATCTATTGGAGGAGTAACAAGAATGAGCTGGCACGGCAACACTATTCACT 240

Qy    2780 gatcaggaactgtccgccgaagattttgacagacaatctgggtttggcggttggccacgc 2839
           |||||||||||||| |||||||| ||||||||||||||||||||||||||||||||||||
Db     241 GATCAGGAACTGTCCNCCGAAGATTNTGACAGACAATCTGGGTTTGGCGGTTGGCCACGC 300

Qy    2840 gttgtgtgctcgcaagatttgcatagatgaccgagattccccgaaagtcagtcaatacgt 2899
           |||||||||||||||||| |||||||||||| |||||||||||||||| ||||||||||
Db     301 GTTGTGTGCTCGCAAGATTTGNATAGATGACCGANATTCCCCGAAAGTCAGNCAATACGT 360
```

```
RESULT   4
C11998
LOCUS       C11998         378 bp    mRNA            EST       28-DEC-1998
DEFINITION  C11998 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk144h7 5', mRNA sequence.
ACCESSION   C11998
VERSION     C11998.1  GI:1559551
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 378)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A. and Nishigaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
```

```
COMMENT     On Nov 2, 1998 this sequence version replaced gi:3828838.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .378
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /db_xref="taxon:6239"
                     /clone="yk144h7"
                     /clone_lib="Yuji Kohara unpublished cDNA"
                     /sex="hermaphrodite, male"
                     /tissue_type="whole animal"
                     /dev_stage="varied"
BASE COUNT       93 a     96 c     90 g     89 t     10 others
ORIGIN

  Query Match             10.3%;  Score 352.8;  DB 80;  Length 378;
  Best Local Similarity   96.6%;  Pred. No. 2.9e-93;
  Matches  365;  Conservative    0;  Mismatches   12;  Indels    1;  Gaps    1;

Qy   2757 ctggcacggcaacactattcactgatcaggaactgtccgccgaagattttgacagacaat 2816
          ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db      1 CTGGCACGGCAACACTATTCACTGATCAGGAACTGTCCGCCGAAGATNTTGACAGACAAT 60

Qy   2817 ctgggtttggcggttggccacgcgttgtgtgctcgcaagatttgcatagatgaccgag-a 2875
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     61 CTGGGTTTGGCGGTTGGCCACGCGTTGTGTGCTCGCAAGATTTGNATAGATGACCGAGCA 120

Qy   2876 ttccccgaaagtcagtcaatacgtgtgcattcacacaaagaagtcgctcgaatccctccg 2935
          |||||||||||||||||||||||||||||| |||||||||||||||||||| ||||||||
Db    121 TTCCCCGAAAGTCAGTCAATACGTGTGCATNCACACAAAGAAGTCGCTCGACTCCCTCCG 180

Qy   2936 actattctccacatcatcgcgagcatcaggtgtggtgtctggaattcaggaaggtacacg 2995
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    181 NCTATTCTCCACATCATCGNGAGCATCAGGTGTGGTGTCTGGAATTCAGGAAGGTACACG 240

Qy   2996 ccgaatggcctacgaatggattatgaactcgctgctcgacgcgtggcgttccaatctatt 3055
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    241 CCGAATGGCCTACGAATGGATTATGANCTCGCTGCTCGACGCGTGGCGTTCCAATCTATT 300

Qy   3056 cgcatcgaaaccctactggacacaaagcttcaagggacaatccacgtttagtacgcttta 3115
          |||||||||||||||| |||||| |||||||||||||| ||||||||||||||||||||
Db    301 CGCATCGAAACCCTACTNGACACANAGCTTCAAGGGACANTCCACGTTTAGTACGCTTTA 360

Qy   3116 tcaagaggcgtataatca 3133
          ||| |||| |||||||||
Db    361 TCANGAGGTGTATAATCA 378


RESULT   5
D67853
LOCUS       D67853        360 bp    mRNA            EST       07-DEC-1995
DEFINITION  CELK087G4F Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk87g4 5', mRNA sequence.
ACCESSION   D67853
VERSION     D67853.1  GI:1105632
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
            Tabara,H.
  TITLE     Toward an expression map of the C.elegans genome
  JOURNAL   Unpublished (1994)
COMMENT     Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /db_xref="taxon:6239"
                     /clone="yk87g4"
                     /clone_lib="Yuji Kohara unpublished cDNA"
                     /sex="hermaphrodite, male"
                     /tissue_type="whole animal"
                     /dev_stage="varied"
BASE COUNT       98 a     86 c     85 g     88 t      3 others
ORIGIN

  Query Match             10.3%;  Score 350.6;  DB 81;  Length 360;
  Best Local Similarity   98.1%;  Pred. No. 1.3e-92;
  Matches  353;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy   1303 gattaccgtacccggaaccaattcaatacactgaatactctgcccgatcacccgtcgaat 1362
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GATTACCGTACCCGGAACCAATTCAATACACTGAATACTCTGCCCGATCACCCGTCGAAT 60

Qy   1363 catctccttctccacctagaaatgagagaaaacgatcacgaatgagcacaacgactccta 1422
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CATCTCCTTCTCCACCTAGAAATGAGAGAAAACGATCACGAATGAGCACAACGACTCCTA 120

Qy   1423 tgaagaatggaactagagatggatcttcgaaagttacccttttgcgatgctcctagcag 1482
          |||||||||||||||||||||||||||||||||||| | ||||||||||||||| ||
Db    121 TGAAGAATGGAACTAGAGATGGATCTTCGAAAGTTACCCTNTGTGCGATGCTCCTAGNAG 180

Qy   1483 ttctgatttttaatccgattggattgctcgctggaagtgcgatattctcaaaagccgctg 1542
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    181 TTCTGATTTTTAATCCGATTGGATTGCTCGCTGGAAGTGCGATATTCTCAAAAGCCGNTG 240

Qy   1543 cagaagctccgattgcctccccgttcgagcatggaagagtgattgatgacccggatggaa 1602
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    241 GAGAAGCTCCGATTGCCTCCCCGCTCGAGCATGGAAGAGTGATTGATGACCCGGATGGAA 300

Qy   1603 ctagcactcggacgcttttctgggaagggagtatcatcaatatgagctatgtctgggtgt 1662
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    301 CTAGCACTCGGACGCTTCTCTGGGAAGGGAGTATCATCAATATGAGCTATGTCTGGGTGT 360


RESULT   6
D34815
LOCUS       D34815        360 bp    mRNA            EST       08-AUG-1994
DEFINITION  CELK012HZF Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk12h12 5', mRNA sequence.
ACCESSION   D34815
VERSION     D34815.1  GI:526331
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
            Tabara,H.
  TITLE     Toward an expression map of the C.elegans genome
  JOURNAL   Unpublished (1994)
COMMENT     On Apr 7, 1998 this sequence version replaced gi:3035947.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
```

```
           Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /db_xref="taxon:6239"
                     /clone="yk12h12"
                     /clone_lib="Yuji Kohara unpublished cDNA"
                     /sex="hermaphrodite, male"
                     /tissue_type="whole animal"
                     /dev_stage="varied"
BASE COUNT      119 a    106 c     58 g     74 t      3 others
ORIGIN


  Query Match           10.2%;  Score 347.4;  DB 81;  Length 360;
  Best Local Similarity   97.5%;  Pred. No. 1.1e-91;
  Matches  351;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy    545 gaatttcactcatgcaatgctatcttcaccacaccatacctcaatgacttctcaaccata 604
          |||||||||||||||||||||||||| |||||||||||||||||||||| ||||||||||
Db      1 GAATTTCACTCATGCAATGCTATCTCCACCACACCATACCTCAATGACTCCTCAACCATA 60

Qy    605 tacagaagccatgggacatatcaacgggtacatgtctccatacgaccaagctcaaggccc 664
          ||||||||| ||||  |||||||||||||||||||||||||||||||||||||||||| ||
Db     61 TACAGAAGCAATGGAGCATATCAACGGGTACATGTCTCCATACGACCAAGCTCAAGGNCC 120

Qy    665 atcaggaccatcatattactcacaacaccatcaatctccaccacctcatcaccaccatca 724
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATCAGGACCATCATATTACTCACAACACCATCAATCTCCACCACCTCATCACCACCATCA 180

Qy    725 ccacccgatgccaaaaatccatgagaaccctgaacaagtggcatctccatcgattgaaga 784
          |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    181 CCACCCGATGCCAAAAATCCATGAGAACCCTGAACAAGTGGNATCTCCATCGATTGAAGA 240

Qy    785 tgctccagagacgaaaccaactcatttggttgaaccacaaagtccaaaaagcccgcagaa 844
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    241 TGCTCCAGAGACGAAACCAACTCATTTGGNTGAACCACAAAGTCCAAAAAGCCCGCAGAA 300

Qy    845 tatgaaagaggagcttcttcggttactagttaacatgtctccgagtgaagttgaacggtt 904
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    301 TATGAAAGAGGAGCTTCTTCGGTTACTAGTTAACATGTCTCCGAGTGAAGTTGGACGGTT 360


RESULT   7
C63905
LOCUS       C63905        360 bp    mRNA            EST       22-SEP-1997
DEFINITION  C63905 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk349h3 5', mRNA sequence.
ACCESSION   C63905
VERSION     C63905.1  GI:2422610
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A. and Nishigaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
COMMENT     On Apr 7, 1998 this sequence version replaced gi:3035983.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /db_xref="taxon:6239"
                     /clone="yk349h3"
                     /clone_lib="Yuji Kohara unpublished cDNA"
                     /sex="hermaphrodite, male"
                     /tissue_type="whole animal"
                     /dev_stage="varied"
BASE COUNT      121 a     92 c     73 g     66 t      8 others
ORIGIN


  Query Match           9.9%;  Score 340;  DB 81;  Length 360;
  Best Local Similarity   97.5%;  Pred. No. 1.8e-89;
  Matches  351;  Conservative    0;  Mismatches    8;  Indels    1;  Gaps    1;

Qy    642 ccatacgaccaagctcaaggcccatcaggaccatcatattactcacaacaccatcaatct 701
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CCATACGACCAANCTCAAGGCCCATCAGGACCATCATATTACTCACAACACCATCAATCT 60

Qy    702 ccaccacctcatcaccaccatcaccacccgatgccaaaaatccatgagaaccctg-aaca 760
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||| ||||
Db     61 CCACCACCTCNTCACCACCATCACCACCCGATGCCAAAAATCCATGAGAACCCTGTAACA 120

Qy    761 agtggcatctccatcgattgaagatgctccagagacgaaaccaactcatttggttgaacc 820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGTGGCATCTCCATCGATTGAAGATGCTCCAGAGACGAAACCAACTCATTTGGTTGAACC 180

Qy    821 acaaagtccaaaaagcccgcagaatatgaaagaggagcttcttcggttactagttaacat 880
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    181 ACAAAGTCCAAAAAGCCCGCAGAATATNAAAGAGGAGCTTCTTCGGTTACTAGTTAACAT 240

Qy    881 gtctccgagtgaagttgaacggttaaagaataaaaaatcaggagcatgttcagcgacgaa 940
          |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    241 GTCTCCGAGTGAAGTTGAACGGTTAAAGAATAAAAAATCNGGAGCATGTTCAGCGACGAA 300

Qy    941 tgggccatcgaggagtaaggagaaggcggcgaagattgtgattcaggagacagcggaagg 1000
          |||||||||||||||| |||||| |||||||| |||||||||||||||||||||||||| |
Db    301 TGGGCCATCGAGGAGTANGGAGAANGCGGCGAANATTGTGATTCAGGAGACAGCGGAANG 360


RESULT   8
D64806/c
LOCUS       D64806        329 bp    mRNA            EST       13-DEC-1995
DEFINITION  CELK012GZR Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk12g12 3', mRNA sequence.
ACCESSION   D64806
VERSION     D64806.1  GI:1115082
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 329)
  AUTHORS   Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
            Tabara,H.
  TITLE     Toward an expression map of the C.elegans genome
  JOURNAL   Unpublished (1994)
COMMENT     On Jun 22, 1998 this sequence version replaced gi:3246664.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .329
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /db_xref="taxon:6239"
                     /clone="yk12g12"
                     /clone_lib="Yuji Kohara unpublished cDNA"
                     /sex="hermaphrodite, male"
                     /tissue_type="whole animal"
```

```
                     /dev_stage="varied"
BASE COUNT       81 a     76 c     76 g     95 t      1 others
ORIGIN


  Query Match             9.5%;  Score 326;  DB 81;  Length 329;
  Best Local Similarity   99.1%;  Pred. No. 2.4e-85;
  Matches  326;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy   3066 ccctactggacacaaagcttcaagggacaatccacgtttagtacgctttatcaagaggcg 3125
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    329 CCCTACTGGACACAAAGCTTCAAGGGACAATCCACGTTTAGTACGCTTTATCAAGAGGCG 270

Qy   3126 tataatcattatgcgattattaatgggacaaggggagattgttggagactatttgtctac 3185
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    269 TATAATCATTATGCGATTATTAATGGGACAAGGGGAGATTGTTGGAGACTATTTGTCTAC 210

Qy   3186 gagctcacgtgccgaatgctcaacggagccaacccacaagccacgtggtcaggcgyccga 3245
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db    209 GAGCTCACGTGCCGAATGCTCAACGGAGCCAACCCACAAGCCACGTGGTCAGGCGTCCGA 150

Qy   3246 cgcgttcgatctacaaaaatggacgcggtccgaggaagagtgagcatgcgacgctcggct 3305
          ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    149 CGCGTTCGATCTACACAAATGGACGCGGTCCGAGGAAGAGTGAGCATGCGACGCTCGGCT 90

Qy   3306 caaccggacgcatttcatcttcatacactggttaaactacatacttctatggatctttga 3365
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     89 CAACCGGACGCATTTCATCTTCATACACTGGTTAAACTACATACTTCTATGGATCTTTGA 30

Qy   3366 attgaacaaaaaatgattttattcagaat 3394
          ||||||||| ||||||||||||||||||
Db     29 ATTGAACAANAAATGATTTTATTCAGAAT 1


RESULT   9
D64735/c
LOCUS       D64735         328 bp    mRNA            EST       13-DEC-1995
DEFINITION  CELK087G4R Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk87g4 3', mRNA sequence.
ACCESSION   D64735
VERSION     D64735.1  GI:1116421
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 328)
  AUTHORS   Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
            Tabara,H.
  TITLE     Toward an expression map of the C.elegans genome
  JOURNAL   Unpublished (1994)
COMMENT     On Oct 6, 1998 this sequence version replaced gi:3705561.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .328
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /db_xref="taxon:6239"
                     /clone="yk87g4"
                     /clone_lib="Yuji Kohara unpublished cDNA"
                     /sex="hermaphrodite, male"
                     /tissue_type="whole animal"
                     /dev_stage="varied"
BASE COUNT       81 a     75 c     75 g     91 t      6 others
ORIGIN
```

```
  Query Match             8.9%;  Score 305.8;  DB 81;  Length 328;
  Best Local Similarity   97.0%;  Pred. No. 2.2e-79;
  Matches  318;  Conservative    1;  Mismatches    8;  Indels    1;  Gaps    1;

Qy   3068 ctactggacacaaagcttcaagggacaatccacgtttagtacgctttatcaagaggcgta 3127
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    328 CTACTGGACACAAAGCTTCAAGGGNCAATCCACGTTTAGTACGCTTTATCAAGAGGCGTA 269

Qy   3128 taatcattatgcgattattaatgggacaaggggagattgttggagactatttgtctacga 3187
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    268 TAATCATTATGCGATTATTAATGGGACAAGGGGAGATTGTTGGAGACTATTTGTCTNCGA 209

Qy   3188 gctcacgtgccgaatgctcaacggagccaacccacaagccacgtggtcaggcgyccgacg 3247
          ||||||||| |||||||||| ||||||||||||||||||||||||||||||:||||||
Db    208 GCTCACGTGNCGAATGCTCACCGGAGCCAACCCACAAGCCACGTGGTCAGGCGTCCGACG 149

Qy   3248 cgttcgatctacaaaaatggacgcggtccgaggaagagtgagcatgcgacgctcggctca 3307
          |||||||||||||| |||||||||||||||||||||||||||||| ||||||||||||||
Db    148 CGTTCGATCTACACAAATGGACGCGGTCCGAGGAAGAGTGAGCATNCGACGCTCGGCTCA 89

Qy   3308 accggacgcatttcatcttcatacactggttaaactacatactt-ctatggatctttgaa 3366
          |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db     88 ACCGGACGCATTTCATCTTCATACACTGGTTAAACTACATACTTNCTATGGATCTTTGAA 29

Qy   3367 ttgaacaaaaaatgattttattcagaat 3394
          |||||| | ||||||||||||||||||
Db     28 TTGAACNACAAATGATTTTATTCAGAAT 1


RESULT  10
C10363/c
LOCUS       C10363         252 bp    mRNA            EST       28-DEC-1998
DEFINITION  C10363 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk144h7 3', mRNA sequence.
ACCESSION   C10363
VERSION     C10363.1  GI:1535434
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 252)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A. and Nishigaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
COMMENT     On Jul 8, 1999 this sequence version replaced gi:5422657.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .252
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /db_xref="taxon:6239"
                     /clone="yk144h7"
                     /clone_lib="Yuji Kohara unpublished cDNA"
                     /sex="hermaphrodite, male"
                     /tissue_type="whole animal"
                     /dev_stage="varied"
BASE COUNT       57 a     63 c     62 g     66 t      4 others
ORIGIN


  Query Match             6.7%;  Score 229.2;  DB 80;  Length 252;
  Best Local Similarity   96.0%;  Pred. No. 9.6e-57;
  Matches  242;  Conservative    1;  Mismatches    8;  Indels    1;  Gaps    1;

Qy   3093 caatccacgtttagtacgctttatcaagaggcgtataatcattatgcgattattaatggg 3152
```

```
          |||||||||||||||||||||||||| |||||||||||||||||| ||||||||||||||
Db    252 CAATCCACGTTTAGTACGCTTTATCNAGAGGCGTATAATCATTATACGATTATTAATGGG 193

Qy   3153 acaaggggagattgttggagactatttgtctacgagctcacgtgccgaatgctcaacgga 3212
          |||||||||| |||||||||||||||||| |||||||||| |||||||||| ||||
Db    192 ACAAGGGGAGANTGTTGGAGACTATTTGTCTNCGAGCTCACGTGNCGAATGCTCACCGGA 133

Qy   3213 gccaacccacaagccacgtggtcaggcgyccgacgcgttcgatctacaaaaatggacgcg 3272
          ||||||||||||||||||||||||||:|||||||||||||||| ||||||||||
Db    132 GCCAACCCACAAGCCACGTGGTCAGGCGTCCGACGCGTTCGATCTACACAAATGGACGCG 73

Qy   3273 gt-ccgaggaagagtgagcatgcgacgctcggctcaaccggacgcatttcatcttcatac 3331
          || |||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     72 GTGCCGAGGAAGAGTGAGCATACGACGCTCGGCTCAACCGGACGCATTTCATCTTCATAC 13

Qy   3332 actggttaaact 3343
          ||||||||||||
Db     12 ACTGGTTAAACT 1


RESULT  11
AW115144
LOCUS       AW115144      337 bp    mRNA            EST       20-OCT-1999
DEFINITION  rs64a09.y1 Sommer Pristionchus Pristionchus pacificus cDNA 5'
            similar to WP:Y47D3B.7 CE19180 HELIX-LOOP-HELIX DNA-BINDING DOMAIN
            ;, mRNA sequence.
ACCESSION   AW115144
VERSION     AW115144.1  GI:6081482
KEYWORDS    EST.
SOURCE      Pristionchus pacificus.
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 337)
  AUTHORS   McCarter,J., Clifton,S., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M.,
            Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T.,
            Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S.,
            Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
            Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Apr 7, 1998 this sequence version replaced gi:3034820.
            Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            The library was constructed by Dr. Ralf Sommer DNA Sequencing by:
            Washington University Genome Sequencing Center
             Contact Dr. Ralf Sommer (ralf.sommer@tuebingen.mpg.de) for
            information about this clone.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 331.
FEATURES             Location/Qualifiers
     source          1. .337
                     /organism="Pristionchus pacificus"
                     /strain="PS 312"
                     /db_xref="taxon:54126"
                     /clone_lib="Sommer Pristionchus"
                     /sex="predominantly hermaphroditic"
                     /dev_stage="mixed stages (embryo to adult)"
                     /lab_host="not applicable (host cell line)"
                     /note="Vector: Uni-ZAP XR Vector (Stratagene); Site_1: 5'
                     EcoRI; Site_2: 3' XhoI; 1st strand cDNA was primed with a
                     XhoI - oligo(dT) primer. Double-stranded cDNA was ligated
                     to EcoRI adaptors digested with XhoI and cloned into XhoI
                     and EcoRI sites. Primary complexity of the library was 10
                     in the 7th. The library went through one round of
                     amplification."
BASE COUNT       71 a    107 c     80 g     79 t
ORIGIN


  Query Match           1.8%;  Score 62.8;  DB 64;  Length 337;
  Best Local Similarity   52.3%;  Pred. No. 1.9e-07;
  Matches  139;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy   2107 ccctctgtgcagtaaaccttgctgaagctgccggagcatcaaacgacggacttccacgcg 2166
          |||||   || |||||  |  | ||  |||| ||  ||||   ||| |||||||| | |
Db      1 CCCTCACAGCCGTAAATATCTCGGAGTCTGCAGGTTCATCCGTCGATGGACTTCCGCACA 60

Qy   2167 ccgtcatggctcagatctacatttctgcatccatccaatgccgtttggctcttccgaaccc 2226
            | |   ||  | ||| ||||    ||  |  | |  |||||  |  | ||||| ||
Db     61 TTGACCACGCCAACATCCACATGGGAGCTGCTCTGCGTTGCCGCATCACCCTTCCCAAGT 120

Qy   2227 tactcgcaccattcttctcgggatacttttacgaagagctcgaaggcacgtgcgtcgag 2286
          |  |||| ||| |  || |    ||||| ||     |   | | || ||||||| |
Db    121 TCTTCGCCCCAGTGCTCGCCAAGTACTTCTTCACGCGCAGTAGGAGACACGTGAGGAAGA 180

Qy   2287 ctccggagcactcggtgtcccatttgttatggatcttccatccagcgacaagaaagttca 2346
            |                | ||    |    |  ||||| |||||  |    | || |||
Db    181 TGTCACCAGGAAGTTCGCCCGCCCTCGGCTTCATCTTTCATCCGATGTGCCGTAAATTCG 240

Qy   2347 tgtcagatgcgaaaaggttggagcat 2372
          | || ||| | |   | |||  | ||
Db    241 TCTCTGATCCTACCCGATTGAGGAAT 266


RESULT  12
AW701427
LOCUS       AW701427      506 bp    mRNA            EST       18-APR-2000
DEFINITION  uq86a04.y1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:2937966 5'
            similar to gb:U00968 STEROL REGULATORY ELEMENT BINDING PROTEIN 1
            (HUMAN);, mRNA sequence.
ACCESSION   AW701427
VERSION     AW701427.1  GI:7585558
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 506)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 6, 2000 this sequence version replaced gi:6677089.
            Other_ESTs: uq86a04.x1
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml

            MGI:1050386
            Seq primer: -40RP from Gibco
            High quality sequence stop: 449.
FEATURES             Location/Qualifiers
     source          1. .506
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2937966"
                     /clone_lib="NCI_CGAP_Lu33"
                     /tissue_type="pooled lung tumors"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st
```

```
                strand cDNA was prepared from mRNA obtained from pooled
                lung tumors with a Not I - oligo(dT) primer [5'
                TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCTGTTTTTTTTTTTTTTTT 3'].
                Double-stranded cDNA was ligated to Eco RI adaptors
                (Pharmacia),  digested with Not I and cloned into the Not
                I and Eco RI sites of  the modified pT7T3 vector. Library
                went through one round of normalization, and was
                constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      130 a    146 c    142 g     88 t
ORIGIN


  Query Match             1.7%;  Score 58.4;  DB 80;  Length 506;
  Best Local Similarity   52.0%;  Pred. No. 4.5e-06;
  Matches  131;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy   1086 gagcgtcgtacggcacacaatctcatcgaaaagaagtatagatgctcaataaatgatcga 1145
          |||   || || || ||||||  ||| || |||   ||  | | ||| || |||||
Db     49 GAGAAGCGCACAGCCCACAATGCCATTGAGAAGCGCTACCGGTCCTCTATCAATGACAAG 108

Qy   1146 attcaacagctgaaagtacttttgtgtggggatgaagctaagctttcaaaatcggcaaca 1205
          |||    |||| ||||  ||  ||   ||   |||||| |||||    ||||| ||
Db    109 ATTGTGGAGCTCAAAGACCTGGTGGTGGGCACTGAAGCAAAGCTGAATAAATCTGCTGTC 168

Qy   1206 ctacgacgggctattgaacatatcgaggaggttgaacacgagaatcaggtgttgaagcat 1265
          | ||   ||| || ||  | |||        |  | |||   || |||  | | |||||
Db    169 TTGCGCAAGGCCATCGACTACATCCGCTTCTTGCAGCACAGCAACCAGAAGCTCAAGCAG 228

Qy   1266 catgttgaacaaatgagaaagacactgcagaataatcgattaccgtacccggaaccaatt 1325
          |           |  |||   |||   | |  ||   | |   | ||| ||   ||  |
Db    229 GAGAACCTGACCCTACGAAGTGCACACAAAAGCAAATCACTGAAGGACCTGGTGTCAGCT 288

Qy   1326 caatacactgaa 1337
             || || |
Db    289 TGTGGCAGTGGA 300


RESULT  13
CNS0073W/c
LOCUS       CNS0073W      922 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR14D09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066784
VERSION     AL066784.1  GI:4945247
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 922)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .922
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone_lib="RPCI-98"
                      /clone="BACR14D09"
                      /note="end : TET3"
BASE COUNT      223 a     95 c    109 g    221 t    274 others
ORIGIN


  Query Match             1.6%;  Score 55.4;  DB 122;  Length 922;
  Best Local Similarity   20.9%;  Pred. No. 4.4e-05;
  Matches   80;  Conservative  139;  Mismatches  162;  Indels    1;  Gaps    1;

Qy    466 ctcaacaactcgccgcgccagcgatgactccacatcaggcagcgtcgctttttgttaata 525
          | :::|::| | || | ::: :    :| ||:|: :   :  :  : :         :: |
Db    922 CMMMMCMMCCCMCCCCCMMMCMMCMMMCMCCMCMMMCMMMCMMMMMCMCMCMMMMMMMA 863

Qy    526 ctaatggaattgatcaaaagaatttcactcatgcaatgctatcttcaccacaccatacct 585
          |  |   |:   : |:::: ::   :|: |:  |::  : : :  ::: ::|:|| : :
Db    862 CACAMMAAMNMMMACMMMMCMMACMMAMCCMMACMMMAMAMMMMMMMMAMMAMCACMAMM 803

Qy    586 caatgacttctcaaccatatacagaagccatgggacatatcaacgggtacatgtctccat 645
          :|   :|  : : :::: : :|:  : |:     ::| | :|:|   ::    | |::
Db    802 MACACMCAMMMMCMMMMMMMMCMMCMMCMCCACMMMACACMAMCCMMCMMCMMACMCMMA 743

Qy    646 acgaccaagctcaaggcccatcaggaccatcatattactcacaacaccatcaatctccac 705
          |: |:::: | ::|   ::: :|  |::| :: |  :: ::| :| : |  :| : |::|
Db    742 AMAAMMMMACAMMAMAAMMMMMAMAAMMAAMMAAMMMMAMMCCMCCMAAMAMAMMACMMC 683

Qy    706 cacctcatcac-caccatcaccacccgatgccaaaaatccatgagaaccctgaacaagtg 764
          ::|  :: ::| ::::: :: :::|| |   :: : | :::  | :|:::   :::|
Db    682 MMCAMMMMMMCAMMMMAMMAMMMCCCAACAMMCMCACMMMCMAAMAMMMMACMMMAMMA 623

Qy    765 gcatctccatcgattgaagatgctccagagacgaaaccaactcatttggttgaaccacaa 824
           :: | | | : |   |: | | :|: |:|| |::::| ::        ||::|::|
Db    622 AMMACMCAAMMCACAAAMMACMCMMCMCASACACAMVMRAAMMMMCCCASAMAAMMAMMA 563

Qy    825 agtccaaaaagcccgcagaata 846
          |  |  :|:  | :  |:|| |
Db    562 AMACACMAMCACASAAARAAMA 541


RESULT  14
CNS00LO0
LOCUS       CNS00LO0      1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR32D23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL068607
VERSION     AL068607.1  GI:4958689
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
```

isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR32D23"
/note="end : TET3"
BASE COUNT 155 a 166 c 7 g 284 t 489 others
ORIGIN

Query Match 1.5%; Score 52.8; DB 122; Length 1101;
Best Local Similarity 18.0%; Pred. No. 0.00027;
Matches 81; Conservative 171; Mismatches 197; Indels 0; Gaps 0;

```
Qy    393 cttaacccctacaaacatctccaccaagtggaggttacccccggcagatgcctacaga 452
          ||  ::| |:  |:::   : ::: |::   :   :: : ||:  :  :  : : :: |
Db    497 CTCYVMCACYATCMMHACCMMHMMACMMACTMMMMCYMAYTCCHTYMCHMMCMTHCMMTA 556

Qy    453 cctctatcacttgctcaacaactcgccgcgccagcgatgactccacatcaggcagcgtcg 512
          :| :| |  | : :||:::| | : |: :  |: | |:  : :   :|::        |:
Db    557 MCMMTCTATCMHCMTCMHMAYCMYCCMCYTTCMCCMAHHTMAMACAMTMMAHACHTHTHW 616

Qy    513 ctttttgttaatactaatggaattgatcaaaagaatttcactcatgcaatgctatcttca 572
          | | ||  |:  :: |:|  :    : : :||  :||: :: ::  :|:  |::|    :
Db    617 CMTMTTMATMTCMMCAHTMAHTMMTMAMCHAAHNWTTWTMMMMHMAYAHCACWWTTMCAM 676

Qy    573 ccacaccatacctcaatgacttctcaaccatatacagaagccatgggacatatcaacggg 632
          ::::   : ::: :::  ::| :||| ::: | : :  :   ::   ||: : :::|  :
Db    677 MMWMCAAMCMHMMMMMMAMHTAMTCACMMMAAMMAMMTMWTTWWAMMACMRMAMMMCCCS 736

Qy    633 tacatgtctccatacgaccaagctcaaggcccatcaggaccatcatattactcacaacac 692
           !::   : : :::  :  :: : |::  :::: ||  :  |  : |  :|  :: | :::
Db    737 CAMMCMAMAMAHWMACMAWHMCMMCMMMCMMMMMCAMAMWTACAMCAAMAAAMWWAYMMH 796

Qy    693 catcaatctccaccacctcatcaccaccatcaccacccgatgccaaaaatccatgagaac 752
          || ::: : :::: :|  :| |:|::|:  |::|:||| :  : |: :: ::|    : :
Db    797 CACMMMMYCMMMMAHCTMYACCMCMMCYCACMYCWCCCMMMCMWAMTMMAMMAMTTTHCH 856

Qy    753 cctgaacaagtggcatctccatcgattgaagatgctccagagacgaaaccaactcatttg 812
          | : :  |  |:|: ||: : :  ::    |::: : :  ::|:  | |  :
Db    857 TTTCMYHTTTTTYCMTMCCCMMMMMMAHMMHCMMWTHMWHWAHAWTHMCMCCCCCTMMWH 916

Qy    813 gttgaaccacaaagtccaaaaagcccgca 841
           :| ::   ::::  | |:::: |:: ::
Db    917 AHTWWHWWTHHHHWMCAAMMMMACHMAMM 945

RESULT 15
CNS0100X/c
LOCUS CNS0100X 1101 bp DNA GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL098379
VERSION AL098379.1 GI:5609990
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/plasmid="pBeloBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN03G04"
/note="end : SP6"
BASE COUNT 195 a 108 c 131 g 161 t 506 others
ORIGIN

Query Match 1.5%; Score 50.4; DB 122; Length 1101;
Best Local Similarity 19.0%; Pred. No. 0.0014;
Matches 139; Conservative 263; Mismatches 323; Indels 6; Gaps 1;

```
Qy   211 cacaggaatattatgatattgatggtcaacgagacgtaagcaccttacactccctgctca 270
          ::::: :|::::: :::: :: | :   : : :: |:: :: : |: | :
Db   1072 MWMWRHTHTHHHRHTHWHHHTWYTMTATTHHKWMWTYNCRYKCHYNMYNYMTCHMTCCMY 1013

Qy   271 accacaacaacgacgacttcttctcaatgcgattttccccgccaaacttttgatctcggcg 330
           :|:| :|: | ::::::: |: : :|: : :|| |: :| :   :||| |
Db   1012 CHCWCCMCWCWHAWMWMWHMWHWAWAWVTHTWTWMWWMCCTCMCTMCMYCYYYCTCTYCY 953

Qy   331 gaggccgtggaccttctctagccgccacccaacaattatctggagaaggtcctgcaagta 390
            :| : |::   : : ::: ::: :|: :|: :: : | :|: :: ::
Db   952 CYYNYCMWAAAHHACAKAAMWAMSHWMYMWCWCHCHTMMMHCTMHACTCAHCHCTMMAYM 893

Qy   391 tgcttaacccctacaaacatctccaccaagtggaggttacccccggcagatgcctaca 450
            :: :::::|: | : :|||: :| :: : :| | | |: |
Db   892 MYMYMHMYYYCYMYCYTYMYYCTCYCHCYCCCCYCYMVHMATMTMCAMTATAMTWATWTA 833

Qy   451 gacctctatcacttgctcaacaactcgccgcgccagcgatgactccacatcaggcagcgt 510
           |::| ::| : : :|:::: ::: | : ::: : | :: : ::::|: |: :
Db   832 CAHHTWWWTAMDHMMMTMMMMCTYHMWCWMYAHHHYHYCTMWHMHTHHWHCHMHCHYMTM 773

Qy   511 cgctttttgttaatactaatggaattgatcaaaagaatttcactcatgcaatgctatctt 570
          : ::::| | || : || || | |||| :| | | | |:| ||
Db   772 HHHHYHTMYTAAAMMAAAAAAAAAAAAMAAAAAAAAMMAAAMATWTATMTTATTTATHTTTT 713

Qy   571 caccacaccatacctcaatgacttctcaaccatatacagaagccatgggacatatcaacg 630
          :| : ::   ::|:|: : | ::: :|: :: ::: :| :::: |:| | ::
Db   712 YATMTYWTWTYHCMTMSHCHCCACMHMYHCHCYMCHYMMHAWMHMWCWMAHACAMAMHAM 653

Qy   631 ggtacatgtctccatacgaccaagctcaaggcccatcaggaccatcatattactcacaac 690
            |:| : :| ::|: |::   : : : : :::: :|::| : : |||
Db   652 MMMAMA------AMMWCCMMCWCWCYMYCTMAHTHMWHWHMWHWWHTHAHHAAHTMAAAC 599

Qy   691 accatcaatctccaccacctcatcaccaccatcaccacccgatgccaaaaatccatgaga 750
          :|:| : :|:|: ::::| |:: : |:|:: : :| | |: ::   :| ::| :
Db   598 WCHAMATHAHTMCHTHMHMTNTTYHTYTCHAYHTYWHCWCWAHAMYKCYHAMMMACAYAM 539

Qy   751 accctgaacaagtggcatctccatcgattgaagatgctccagagacgaaaccaactcatt 810
          |:: : :|::: | :: : | :: |: :: : : : : : |:| |:::   :
Db   538 AMMAHAMAYMMMTMAMHMHAWWCTNHTWBHAWAMHAHMHAMWHMWAHAWAACHHHMWMAM 479

Qy   811 tggttgaaccacaaagtccaaaaagcccgcagaatatgaaagaggagcttcttcggttac 870
          :: ::| | |:| |::|:   : :   : : |:: :   : ::| : : :|
Db   478 AMAHWMMWCAAAAWAWAWAAMWAHAWWHAHTMCWAHAHAMWAHHAYAHAWMTCMCCHCMC 419

Qy   871 tagttaacatgtctccgagtgaagttgaacggttaaagaataaaaaatcaggagcatgtt 930
          : | ||| ::| :| : ||: |:| :|:| ||:| :| || : ||
Db   418 HCMCTCACACNCAYMCACCCCMAAYAAAAMAHAMAAAMHAHACAAAAAMAWAAGTTHTTT 359

Qy   931 cagcgacgaat 941
```
```

```
                |   ::|::|
Db    358 TATATRYGHWT 348


Search completed: September  4, 2000, 00:24:33
Job time: 19848 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2000, 00:48:39 ; Search time 11551.9 Seconds
(without alignments)
582.134 Million cell updates/sec

Title: US-09-332-522B-7
Perfect score: 3768
Sequence: 1 gtttattaagctgcaaatat.....aatattttaacagatttaaa 3768

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: em_fun:*
17: em_hum1:*
18: em_hum2:*
19: em_in:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_sy:*
29: em_un:*
30: em_vi:*
31: gb_htg1:*
32: gb_htg2:*
33: gb_in1:*
34: gb_in2:*
35: em_ba1:*
36: em_ba2:*
37: em_hum3:*
38: em_hum4:*
39: gb_pr4:*
40: gb_htg3:*
41: gb_htg4:*
42: gb_htg5:*
43: gb_htg6:*
44: gb_htg7:*
45: em_htg1:*
46: em_htg2:*
47: em_htg3:*
48: em_hum5:*
49: gb_pl3:*
50: gb_pr5:*
51: gb_htg8:*
52: gb_htg9:*
53: gb_htg10:*
54: gb_htg11:*
55: gb_htg12:*
56: gb_htg13:*
57: gb_htg14:*
58: gb_in3:*
59: gb_htg15:*
60: gb_htg16:*
61: gb_htg17:*
62: em_htg4:*
63: em_htg5:*
64: em_htg6:*
65: em_htg7:*
66: em_hum6:*
67: gb_htg18:*
68: gb_htg19:*
69: gb_htg20:*
70: gb_htg21:*
71: gb_htg22:*
72: gb_htg23:*
73: gb_htg24:*
74: gb_htg25:*
75: gb_htg26:*
76: gb_htg27:*
77: gb_htg28:*
78: gb_htg29:*
79: gb_htg30:*
80: gb_htg31:*
81: gb_vi1:*
82: gb_vi2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	3768	100.0	3768	33	DMU38238	U38238 Drosophila
	2	3053	81.0	44517	42	AC018013	AC018013 Drosophil
c	3	3053	81.0	273521	34	AE003515	AE003515 Drosophil
c	4	853.2	22.6	57873	51	AC010021	AC010021 Drosophil
c	5	714.6	19.0	92677	51	AC009382	AC009382 Drosophil
	6	524.4	13.9	57873	51	AC010021	AC010021 Drosophil
	7	462.8	12.3	92677	51	AC009382	AC009382 Drosophil
	8	166.6	4.4	4154	50	U00968	U00968 Human SREBP
	9	165	4.4	4154	5	AR069298	AR069298 Sequence
	10	165	4.4	4154	5	I22310	I22310 Sequence 37
	11	114.4	3.0	4249	9	HSU02031	U02031 Human stero
	12	114.4	3.0	5197	5	AR069307	AR069307 Sequence
	13	114.4	3.0	5197	5	I22319	I22319 Sequence 53
	14	112	3.0	3816	12	CGU09103	U09103 Cricetulus
	15	105.8	2.8	3115	12	RATADD1A	L16995 Rat add1 mR
	16	90.2	2.4	1209	12	AB017337	AB017337 Mus muscu
	17	88.6	2.4	433	3	AF102873	AF102873 Sus scrof
	18	88.6	2.4	4987	12	GU12330	U12330 Cricetulus
	19	82.6	2.2	1767	12	CGU22819	U22819 Cricetulus
	20	82.6	2.2	1914	12	CGU12329	U12329 Cricetulus
	21	82.6	2.2	4425	12	CGU22818	U22818 Cricetulus
	22	80.8	2.1	220895	10	HS250D10	Z99716 Human DNA s
	23	64	1.7	89299	32	HS17B87	AL035362 Homo sapi
c	24	51.6	1.4	110000	31	PFMAL4P1_2	Continuation

c	25	51.4	1.4	110000	52	AC010182_0
	26	50.8	1.3	1156	12	MMU80889
c	27	50.4	1.3	2126	12	MMU80891
	28	49.4	1.3	477	11	AF020275
c	29	49.4	1.3	95968	33	CEY47D3B
c	30	49.4	1.3	197735	31	CEY66A7
c	31	49.4	1.3	337565	31	CEY47D3
c	32	49	1.3	27694	7	HASMT
	33	48.8	1.3	3071	33	DME131022
	34	48.4	1.3	234	5	AR002264
	35	48.4	1.3	234	5	AR053098
	36	48.4	1.3	1159	39	AF126748
c	37	48.4	1.3	1442	39	AF087653
c	38	48.4	1.3	1472	39	AF126749
c	39	48.2	1.3	5719	12	MMCATS1
c	40	47.8	1.3	137271	50	AC004552
	41	47.6	1.3	720	9	AB007820
	42	47.6	1.3	1410	9	AB009843
	43	47.6	1.3	1490	12	MMU70643
	44	47.6	1.3	1546	12	MMU70642
	45	47.6	1.3	1546	12	MMU70645

AC010182 Homo sapi
U80889 Mus musculu
U80891 Mus musculu
AF020275 Homo sapi
AL031635 Caenorhab
AL022282 Caenorhab
Z98865 Caenorhabdi
D31785 Hansenula w
AJ131022 Drosophil
AR002264 Sequence
AR053098 Sequence
AF126748 Homo sapi
AF087653 Homo sapi
AF126749 Homo sapi
AF051726 Mus muscu
AC004552 Homo sapi
AB007820 Homo sapi
AB009843 Homo sapi
U70643 Mus musculu
U70642 Mus musculu
U70645 Mus musculu

ALIGNMENTS

RESULT 1
DMU38238
LOCUS DMU38238 3768 bp mRNA INV 25-APR-1996
DEFINITION Drosophila melanogaster transcription factor HLH106 mRNA, complete cds.
ACCESSION U38238
VERSION U38238.1 GI:1079655
KEYWORDS .
SOURCE fruit fly strain=Canton S.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 3768)
AUTHORS Theopold,U., Ekengren,S. and Hultmark,D.
TITLE HLH106, a Drosophila transcription factor with similarity to the vertebrate sterol responsive element binding protein
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 93 (3), 1195-1199 (1996)
MEDLINE 96165543
REFERENCE 2 (bases 1 to 3768)
AUTHORS Hultmark,D.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-1995) Dan Hultmark, Molecular Biology, Stockholm University, S-106 91 Stockholm, Sweden

FEATURES Location/Qualifiers
source 1. .3768
/organism="Drosophila melanogaster"
/strain="Canton S"
/db_xref="taxon:7227"
CDS 180. .3521
/note="SREBP homolog; bHLH protein; Method: conceptual translation supplied by author"
/codon_start=1
/function="transcription factor"
/product="HLH106"
/protein_id="AAA97864.1"
/db_xref="GI:1079656"
/translation="MDTTLMNLIDAPLDESMDLFKAEDVFEPFDADLHSDMLDIILND
MDLAPTQMYNMLLDEPRTHTQQTQSVDQQPQSVEQQPHVKSEHSSPVHIKEELHQQQQ
QSPLLVYKPDPLIATSYNCPQQQPTGLLKAAQPTATIHHMDAQRMPPNTAVYPPSLGS
SFVYQSMSPPTSPVESANQNVNVMQPVAATPAPASAPLPQQSYPQPFITYNSKAGMTS
DEAMYLLLQPTVASPTPSPPVAPPPTSTGSRASKVRVAPLAPSPAAMEVQGKVPINRV
QPKVKEVKRSAHNAIERRYRTSINDKINELKNLVVGEQAKLNKSAVLRKSIDKIRDLQ
RQNHDLKAELQRLQRELMARDGSKVKDLLQLGTRPGRASKKRRESSQTFTTDAGLTPP
RSDESDPSLSPMHSDISLPPSPYGGSTASCSSGSSSSNEEPLVVPSSMRGMATHSRLG
LCMFMFAILAVNPFKTFLQRGHYDSNDDLGDMSGQRRILSYDVEGEGFAVWQQSSWIW
LLNFTLMLGCLVKLLVYGDPQLDAQTDAYCQHRQRADFYFSQGQSSQAYAGYLNCLHM
FGLSLPASRLECYLQTTWQFLRFLFHRLWLGRVLSRRSGGLFSNAASRKQALASAREL
ALLFNRLNQLQLTGNGSRGDMNGIMMALFASNMAEVAHNLLTPRETICIHVMTALRMK
RSAPKWLQQFFARYYMSRARQECGRTRATEQTQELRWAFTAYGYRYCATHVFTYDLSD
SGEQDGFFTRLRNPCDPAAHVIKQYREHLLFKSIQCLVGAGHKSGGLPTSSVSGEAEQ
LQQQQHSGTIVSNVLKYTSLLKDTLWADEDERDTNVVWWADVLETAVHWLLGEDTLAE
QLYGRIKQMPTQLQQCGENDHLPKALHAVLRAKMILLKNNGNALDKSLKQLVNILCDE
SSVELQECLTVNRITDAKGIKLLFQLLTCDWLLETRTALWELEHMNMEDDGFYQVPGE
VLEKFQTDLNSLRNIVENIPNAQSRIYLYEAVCRLMAGASPCPTQQLLDRSLRSRNAH
SSIFCGSKDRRQQNFVGGERERASAMYVACKYLPPALLSSPGERAGMLAEAAKTLEKV
GDKRKLKECYQLMKSLGNGIGSVKA"

BASE COUNT 945 a 1030 c 984 g 809 t
ORIGIN

Query Match 100.0%; Score 3768; DB 33; Length 3768;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3768; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 gtttattaagctgcaaatatactcgtgaaaaaaatcaaaacaaccatgaacaacaagtgt 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GTTTATTAAGCTGCAAATATACTCGTGAAAAAAATCAAAACAACCATGAACAACAAGTGT 60

Qy      61 tgcaactattactaactagtcgctagtttaaagcaaagtgcgttgacattaaccagttat 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGCAACTATTACTAACTAGTCGCTAGTTTAAAGCAAAGTGCGTTGACATTAACCAGTTAT 120

Qy     121 ggaaaaacaaaagcacacgtgaactaagaaaacagatagaaggtggtaaagcattcgcaa 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GGAAAAACAAAAGCACACGTGAACTAAGAAAACAGATAGAAGGTGGTAAAGCATTCGCAA 180

Qy     181 tggacacgacactgatgaacttaatagacgctccgctggacgagtccatggatttgttca 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGGACACGACACTGATGAACTTAATAGACGCTCCGCTGGACGAGTCCATGGATTTGTTCA 240

Qy     241 aagcggaggatgtcttcgaaccgttcgacgccgacctgcactcggacatgctggacatca 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AAGCGGAGGATGTCTTCGAACCGTTCGACGCCGACCTGCACTCGGACATGCTGGACATCA 300

Qy     301 tcctcaacgacatggacctggcgccgacgcagatgtacaacatgctgctggacgagcctc 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TCCTCAACGACATGGACCTGGCGCCGACGCAGATGTACAACATGCTGCTGGACGAGCCTC 360

Qy     361 gaacgcatacccagcagacgcagtccgtggatcagcagccgcaatccgtcgagcaacagc 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GAACGCATACCCAGCAGACGCAGTCCGTGGATCAGCAGCCGCAATCCGTCGAGCAACAGC 420

Qy     421 cgcacgtgaaaagcgagcactcttcgccagtgcacatcaaggaggaactgcatcagcagc 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CGCACGTGAAAAGCGAGCACTCTTCGCCAGTGCACATCAAGGAGGAACTGCATCAGCAGC 480

Qy     481 aacaacagtcgccgcttctcgtctacaaaccagatcccctcatagccacaagctacaatt 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AACAACAGTCGCCGCTTCTCGTCTACAAACCAGATCCCCTCATAGCCACAAGCTACAATT 540

Qy     541 gtccccagcaacagccgacgggccttttgaaggccgcccaaccaacagccaccatacatc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GTCCCCAGCAACAGCCGACGGGCCTTTTGAAGGCCGCCCAACCAACAGCCACCATACATC 600

Qy     601 acatggacgcccagcggatgccgccgaacacggcggtgtatcccccatctctgggcagta 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 ACATGGACGCCCAGCGGATGCCGCCGAACACGGCGGTGTATCCCCCATCTCTGGGCAGTA 660

Qy     661 gctttgtctaccagtccatgtccccgcccacgtcgccggtggagtctgcgaaccagaatg 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GCTTTGTCTACCAGTCCATGTCCCCGCCCACGTCGCCGGTGGAGTCTGCGAACCAGAATG 720

Qy     721 tcaatgtcatgcagcccgttgctgcaactcctgctcccgcttctgctcctttgccccagc 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 TCAATGTCATGCAGCCCGTTGCTGCAACTCCTGCTCCCGCTTCTGCTCCTTTGCCCCAGC 780

Qy     781 agtcgtatccgcaacccttcattacgtacaactctaaggccggaatgacttccgatgaag 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 AGTCGTATCCGCAACCCTTCATTACGTACAACTCTAAGGCCGGAATGACTTCCGATGAAG 840
```

```
Qy    841 ccatgtacttgctcttgcagcccacggtagccagtccaaccccatctccacctgtggctc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CCATGTACTTGCTCTTGCAGCCCACGGTAGCCAGTCCAACCCCATCTCCACCTGTGGCTC 900

Qy    901 caccaccgacaagcacaggtagtcgggccagcaaggtgcgagtggcaccactggctccgt 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CACCACCGACAAGCACAGGTAGTCGGGCCAGCAAGGTGCGAGTGGCACCACTGGCTCCGT 960

Qy    961 cacctgccgctatggaagtccagggcaaggtacctatcaaccgggttcaacccaaggtga 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CACCTGCCGCTATGGAAGTCCAGGGCAAGGTACCTATCAACCGGGTTCAACCCAAGGTGA 1020

Qy   1021 aggaagtaaagcgctcggcccacaacgccatcgagcggcgctatcgcacctcaatcaacg 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 AGGAAGTAAAGCGCTCGGCCCACAACGCCATCGAGCGGCGCTATCGCACCTCAATCAACG 1080

Qy   1081 acaagattaacgagttgaagaacttggtagtgggagagcaggccaagctgaacaagtccg 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 ACAAGATTAACGAGTTGAAGAACTTGGTAGTGGGAGAGCAGGCCAAGCTGAACAAGTCCG 1140

Qy   1141 cagtgttgcggaaatccatagacaagattcgggatctgcaacgccagaatcacgatctga 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CAGTGTTGCGGAAATCCATAGACAAGATTCGGGATCTGCAACGCCAGAATCACGATCTGA 1200

Qy   1201 aggcagagttgcagcgcctgcagagggagctaatggcacgcgacggctccaaggtgaagg 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 AGGCAGAGTTGCAGCGCCTGCAGAGGGAGCTAATGGCACGCGACGGCTCCAAGGTGAAGG 1260

Qy   1261 atttacttcagctgggcactcggcctggtagagcatccaagaagcgccgcgagagctcgc 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 ATTTACTTCAGCTGGGCACTCGGCCTGGTAGAGCATCCAAGAAGCGCCGCGAGAGCTCGC 1320

Qy   1321 agacctttaccacggatgccggactgacgccgccacgcagcgatgaatcggatccttcgc 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 AGACCTTTACCACGGATGCCGGACTGACGCCGCCACGCAGCGATGAATCGGATCCTTCGC 1380

Qy   1381 tctcgcccatgcactcggacatctcgttgccgccatcaccctatggtggatccaccgcca 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TCTCGCCCATGCACTCGGACATCTCGTTGCCGCCATCACCCTATGGTGGATCCACCGCCA 1440

Qy   1441 gctgtagcagtggcagcagcagcagcaatgaagaaccactggtggtgcccagctctatgc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GCTGTAGCAGTGGCAGCAGCAGCAGCAATGAAGAACCACTGGTGGTGCCCAGCTCTATGC 1500

Qy   1501 gcggcatggccacccactctcgcctcggactctgcatgtttatgttcgccatcctggccg 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GCGGCATGGCCACCCACTCTCGCCTCGGACTCTGCATGTTTATGTTCGCCATCCTGGCCG 1560

Qy   1561 tcaatcccttcaagacctttctccagcgcggccactatgacagtaatgacgatcttggcg 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TCAATCCCTTCAAGACCTTTCTCCAGCGCGGCCACTATGACAGTAATGACGATCTTGGCG 1620

Qy   1621 acatgagcggtcaaagacgcattctctcttacgacgtggaaggtgaaggttttgctgtct 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 ACATGAGCGGTCAAAGACGCATTCTCTCTTACGACGTGGAAGGTGAAGGTTTTGCTGTCT 1680

Qy   1681 ggcagcagagttcctggatatggctattgaacttcacactgatgcttggatgcttggtga 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 GGCAGCAGAGTTCCTGGATATGGCTATTGAACTTCACACTGATGCTTGGATGCTTGGTGA 1740

Qy   1741 aattgctggtttacggtgatccgcagctggacgcgcaaacggacgcctactgccagcaca 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 AATTGCTGGTTTACGGTGATCCGCAGCTGGACGCGCAAACGGACGCCTACTGCCAGCACA 1800

Qy   1801 ggcagcgggctgacttctatttagccaaggacagtcgtctcaggcctacgccggttacc 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 GGCAGCGGGCTGACTTCTATTTTAGCCAAGGACAGTCGTCTCAGGCCTACGCCGGTTACC 1860

Qy   1861 tcaactgtctgcatatgtttggattaagtctaccggcgtcgcgcttggagtgttacttgc 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 TCAACTGTCTGCATATGTTTGGATTAAGTCTACCGGCGTCGCGCTTGGAGTGTTACTTGC 1920
```

```
Qy   1921 agaccacgtggcagttccttcgttttctttttccatcgcctctggctgggtcgggtgctgt 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 AGACCACGTGGCAGTTCCTTCGTTTTCTTTTCCATCGCCTCTGGCTGGGTCGGGTGCTGT 1980

Qy   1981 cacggcggtccggtgggctgtttagcaacgccgccagcaggaaacaggcgctggcatctg 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 CACGGCGGTCCGGTGGGCTGTTTAGCAACGCCGCCAGCAGGAAACAGGCGCTGGCATCTG 2040

Qy   2041 cacgcgaactggccctgctcttcaaccgactgaatcaattgcaactgactggaaatggaa 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 CACGCGAACTGGCCCTGCTCTTCAACCGACTGAATCAATTGCAACTGACTGGAAATGGAA 2100

Qy   2101 gccgcggtgacatgaacggcattatgatggcactattcgcaagcaacatggctgaagtgg 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 GCCGCGGTGACATGAACGGCATTATGATGGCACTATTCGCAAGCAACATGGCTGAAGTGG 2160

Qy   2161 cgcacaatctactgacaccgcgcgagaccatctgcatccacgtaatgacagcgttgcgaa 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 CGCACAATCTACTGACACCGCGCGAGACCATCTGCATCCACGTAATGACAGCGTTGCGAA 2220

Qy   2221 tgaagcgcagtgccccaaaatggttgcaacagttcttcgcccgatactacatgagccggg 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 TGAAGCGCAGTGCCCCAAAATGGTTGCAACAGTTCTTCGCCCGATACTACATGAGCCGGG 2280

Qy   2281 ctcgtcaagagtgcggtcgcactagggccaccgagcaaacgcaggagctacgttgggcat 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 CTCGTCAAGAGTGCGGTCGCACTAGGGCCACCGAGCAAACGCAGGAGCTACGTTGGGCAT 2340

Qy   2341 tcacagcctatggatatcgctactgcgccacgcacgtcttcacgtacgatctgagcgact 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 TCACAGCCTATGGATATCGCTACTGCGCCACGCACGTCTTCACGTACGATCTGAGCGACT 2400

Qy   2401 ccggcgagcaggatggattcttcacacgtcttaggaatccatgtgatcccgctgcccacg 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 CCGGCGAGCAGGATGGATTCTTCACACGTCTTAGGAATCCATGTGATCCCGCTGCCCACG 2460

Qy   2461 tcattaagcaatatcgagagcatttgctgtttaaatccattcagtgtctggtaggagcgg 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 TCATTAAGCAATATCGAGAGCATTTGCTGTTTAAATCCATTCAGTGTCTGGTAGGAGCGG 2520

Qy   2521 gccacaaatcgggaggcctgcccacatcttctgtcagcggagaggcggaacagttgcagc 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 GCCACAAATCGGGAGGCCTGCCCACATCTTCTGTCAGCGGAGAGGCGGAACAGTTGCAGC 2580

Qy   2581 aacagcagcacagcggcaccattgtcagcaatgttcttaagtacacgtccctccttaagg 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 AACAGCAGCACAGCGGCACCATTGTCAGCAATGTTCTTAAGTACACGTCCCTCCTTAAGG 2640

Qy   2641 acactctctgggctgatgaggatgaagcgggatacaaacgtggtgtggtgggccgatgttt 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 ACACTCTCTGGGCTGATGAGGATGAGCGGGATACAAACGTGGTGTGGTGGGCCGATGTTT 2700

Qy   2701 tggagaccgcagtgcactggctccttggtgaagacacgctggccgagcaattgtacggca 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 TGGAGACCGCAGTGCACTGGCTCCTTGGTGAAGACACGCTGGCCGAGCAATTGTACGGCA 2760

Qy   2761 ggatcaagcaaatgcccacgcagctgcaacagtgcggcgaaaacgatcatctgcccaagg 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 GGATCAAGCAAATGCCCACGCAGCTGCAACAGTGCGGCGAAAACGATCATCTGCCCAAGG 2820

Qy   2821 cgctgcatgctgtgctgcgagctaagatgatcttactaaaaaacaatggcaacgcactgg 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 CGCTGCATGCTGTGCTGCGAGCTAAGATGATCTTACTAAAAAACAATGGCAACGCACTGG 2880

Qy   2881 acaaaagtctcaagcaattggtaaacatcctctgcgatgagtcgagtgtggagctccaag 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 ACAAAAGTCTCAAGCAATTGGTAAACATCCTCTGCGATGAGTCGAGTGTGGAGCTCCAAG 2940

Qy   2941 agtgcttgactgtcaaccggatcaccgacgccaagggtataaagctgcttttccagttgc 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 AGTGCTTGACTGTCAACCGGATCACCGACGCCAAGGGTATAAAGCTGCTTTTCCAGTTGC 3000

Qy   3001 ttacctgcgattggctgctcgaaactaggactgctctgtgggaactggaacacatgaata 3060
```

```
Db    3001 TTACCTGCGATTGGCTGCTCGAAACTAGGACTGCTCTGTGGGAACTGGAACACATGAATA 3060

Qy    3061 tggaggacgatggcttctaccaagtgccaggtgaagtgctcgagaagttccagaccgatt 3120
Db    3061 TGGAGGACGATGGCTTCTACCAAGTGCCAGGTGAAGTGCTCGAGAAGTTCCAGACCGATT 3120

Qy    3121 tgaactcgttgcgcaacattgtggagaatataccgaacgcccaatcgcgcatatatttgt 3180
Db    3121 TGAACTCGTTGCGCAACATTGTGGAGAATATACCGAACGCCCAATCGCGCATATATTTGT 3180

Qy    3181 acgaggcagtttgtcgcctgatggctggagcctcaccgtgtccaacgcaacagctcttgg 3240
Db    3181 ACGAGGCAGTTTGTCGCCTGATGGCTGGAGCCTCACCGTGTCCAACGCAACAGCTCTTGG 3240

Qy    3241 acaggagtctgcgatcacgcaacgcccactcgtccatcttctgcggcagcaaggatcggc 3300
Db    3241 ACAGGAGTCTGCGATCACGCAACGCCCACTCGTCCATCTTCTGCGGCAGCAAGGATCGGC 3300

Qy    3301 ggcagcagaacttcgtgggcggagagcgggaacgggcttcggccatgtacgtggcctgca 3360
Db    3301 GGCAGCAGAACTTCGTGGGCGGAGAGCGGGAACGGGCTTCGGCCATGTACGTGGCCTGCA 3360

Qy    3361 agtatctcccgcctgcgctgctcagctccccgggtgaacgtgctggcatgttagccgagg 3420
Db    3361 AGTATCTCCCGCCTGCGCTGCTCAGCTCCCCGGGTGAACGTGCTGGCATGTTAGCCGAGG 3420

Qy    3421 cggccaagaccctggagaaggtgggcgacaagcgaaagctcaaggagtgctaccagctga 3480
Db    3421 CGGCCAAGACCCTGGAGAAGGTGGGCGACAAGCGAAAGCTCAAGGAGTGCTACCAGCTGA 3480

Qy    3481 tgaagtcgctgggcaacggcattggcagcgtgaaggcttaggatagtagtgaagtacata 3540
Db    3481 TGAAGTCGCTGGGCAACGGCATTGGCAGCGTGAAGGCTTAGGATAGTAGTGAAGTACATA 3540

Qy    3541 ataagtggcacgaacgtggtgtggattttcagcaaatgaataacccgtttgctattcaaaa 3600
Db    3541 ATAAGTGGCACGAACGTGGTGTGGATTTTCAGCAAATGAATACCCGTTTGCTATTCAAAA 3600

Qy    3601 gaattacaaatgcctaggtctttataattacgctattcctctgttttccacgcccggtta 3660
Db    3601 GAATTACAAATGCCTAGGTCTTTATAATTACGCTATTCCTCTGTTTTCCACGCCCGGTTA 3660

Qy    3661 tgcttagattgtaattttaaaattatttaatatggacattttatttgtttattatttacc 3720
Db    3661 TGCTTAGATTGTAATTTTAAAATTATTTAATATGGACATTTTATTTGTTTATTATTTACC 3720

Qy    3721 gtacttgttaaacgtatttataacaataaatattttaacagatttaaa 3768
Db    3721 GTACTTGTTAAACGTATTTATAACAATAAATATTTTAACAGATTTAAA 3768
```

```
RESULT   2
AC018013
LOCUS       AC018013    44517 bp    DNA             HTG       09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC018013
VERSION     AC018013.1  GI:6553177
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 44517)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10213012 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .44517
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    12813 a    9534 c    9565 g   12605 t
ORIGIN
```

```
  Query Match             81.0%;  Score 3053;  DB 42;  Length 44517;
  Best Local Similarity   87.9%;  Pred. No. 0;
  Matches 3594;  Conservative    0;  Mismatches    5;  Indels  488;  Gaps    5;

Qy     170 agcattcgcaatggacacgacactgatgaacttaatagacgctccgctggacgagtccat 229
Db   18925 AGCATTCGCAATGGACACGACACTGATGAACTTAATAGACGCTCCGCTGGACGAGTCCAT 18984

Qy     230 ggatttgttcaaagcggaggatgtcttcgaaccgttcgacgccgacctgcactcggacat 289
Db   18985 GGATTTGTTCAAAGCGGAGGATGTCTTCGAACCGTTCGACGCCGACCTGCACTCGGACAT 19044

Qy     290 gctggacatcatcctcaacgacatggacctggcgccgacgcagatgtacaacatgctgct 349
Db   19045 GCTGGACATCATCCTCAACGACATGGACCTGGCGCCGACGCAGATGTACAACATGCTGCT 19104

Qy     350 ggacgagcctcgaacgcatacccagcagacgcagtccgtggatcagcagccgcaatccgt 409
Db   19105 GGACGAGCCTCGAACGCATACCCAGCAGACGCAGTCCGTGGATCAGCAGCCGCAATCCGT 19164

Qy     410 cgagcaacagccgcacgtgaaaagcgagcactcttcgccagtgcacatcaaggaggaact 469
Db   19165 CGAGCAACAGCCGCACGTGAAAAGCGAGCACTCTTCGCCAGTGCACATCAAGGAGGAACT 19224

Qy     470 gcatcagcagcaacaacagtcgccgcttctcgtctacaaaccagatcccctcatagccac 529
Db   19225 GCATCAGCAGCAACAACAGTCGCCGCTTCTCGTCTACAAACCAGATCCCCTCATAGCCAC 19284

Qy     530 aagctacaattgtccccagcaacagccgacgggccttttgaaggccgcccaaccaacagc 589
Db   19285 AAGCTACAATTGTCCCCAGCAACAGCCGACGGGCCTTTTGAAGGCCGCCCAACCAACAGC 19344

Qy     590 caccatacatcacatggacgcccagcggatgccgccgaacacggcggtgtatcccccatc 649
Db   19345 CACCATACATCACATGGACGCCCAGCGGATGCCGCCGAACACGGCGGTGTATCCCCCATC 19404

Qy     650 tctgggcagtagctttgtctaccagtccatgtccccgcccacgtcgccggtggagtctgc 709
Db   19405 TCTGGGCAGTAGCTTTGTCTACCAGTCCATGTCCCCGCCCACGTCGCCGGTGGAGTCTGC 19464

Qy     710 gaaccagaatgtcaatgtcatgcagcccgttgctgcaactcctgctcccgcttctgctcc 769
Db   19465 GAACCAGAATGTCAATGTCATGCAGCCCGTTGCTGCAACTCCTGCTCCCGCTTCTGCTCC 19524

Qy     770 tttgccccagcagtcgtatccgcaacccttcattacgtacaactctaaggccggaatgac 829
Db   19525 TTTGCCCCAGCAGTCGTATCCGCAACCCTTCATTACGTACAACTCTAAGGCCGGAATGAC 19584

Qy     830 ttccgatgaagccatgtacttgctcttgcagcccacggtagccagtccaaccccatctcc 889
Db   19585 TTCCGATGAAGCCATGTACTTCCTCTTGCAGCCCACGGTAGCCAGTCCAACCCCATCTCC 19644

Qy     890 acctgtggctccaccaccgacaagcacaggtagtcgggccagcaaggtgcgagtggcacc 949
Db   19645 ACCTGTGGCTCCACCACCGACAAGCACAGGTAGTCGGGCCAGCAAGGTGCGAGTGGCACC 19704

Qy     950 actggctccgtcacctgccgctatggaagtccagggcaaggtacctatcaaccgggttca 1009
Db   19705 ACTGGCTCCGTCACCTGCCGCTATGGAAGTCCAGGGCAAGGTACCTATCAACCGGGTTCA 19764

Qy    1010 acccaaggtgaaggaagtaaagcgctcggcccacaacgccatcgagcggcgctatcgcac 1069
```

```
Db  19765 ACCCAAGGTGAAGGAAGTAAAGCGCTCGGCCCACAACGCCATCGAGCGGCGCTATCGCAC 19824

Qy   1070 ctcaatcaacgacaagattaacgagttgaagaacttggtagtgggagagcaggccaagct 1129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19825 CTCAATCAACGACAAGATTAACGAGTTGAAGAACTTGGTAGTGGGAGAGCAGGCCAAGCT 19884

Qy   1130 gaacaagtccgcagtgttgcggaaatccatagacaagattcgggatctgcaacgccagaa 1189
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19885 GAACAAGTCCGCAGTGTTGCGGAAATCCATAGACAAGATTCGGGATCTGCAACGCCAGAA 19944

Qy   1190 tcacgatctgaaggcagagttgcagcgcctgcagagggagctaatggcacgcgacggctc 1249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19945 TCACGATCTGAAGGCAGAGTTGCAGCGCCTGCAGAGGGAGCTAATGGCACGCGACGGCTC 20004

Qy   1250 caaggtgaaggatttacttcagctgggcactcggcctggtagagcatccaagaagcgccg 1309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20005 CAAGGTGAAGGATTTACTTCAGCTGGGCACTCGGCCTGGTAGAGCATCCAAGAAGCGCCG 20064

Qy   1310 cgagagctcgcagacctttaccacggatgccggactgacgccgccacgcagcgatgaatc 1369
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20065 CGAGAGCTCGCAGACCTTTACCACGGATGCCGGACTGACGCCGCCACGCAGCGATGAATC 20124

Qy   1370 ggatccttcgctctcgcccatgcactcggacatctcgttgccgccatcaccctatggtgg 1429
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20125 GGATCCTTCGCTCTCGCCCATGCACTCGGACATCTCGTTGCCGCCATCACCCTATGGTGG 20184

Qy   1430 atccaccgccagctgtagcagtggcagcagcagcagcaatgaagaaccactggtggtgcc 1489
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20185 ATCCACCGCCAGCTGTAGCAGTGGCAGCAGCAGCAATGAAGAACCACTGGTGGTGCC 20244

Qy   1490 cagctctatgcgcggcatggccacccactctcgcctcggactctgcatgtttatgttcgc 1549
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20245 CAGCTCTATGCGCGGCATGGCCACCCACTCTCGCCTCGGACTCTGCATGTTTATGTTCGC 20304

Qy   1550 catcctggccgtcaatcccttcaagacctttctccagcgcggccactatgacagtaatga 1609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20305 CATCCTGGCCGTCAATCCCTTCAAGACCTTTCTCCAGCGCGGCCACTATGACAGTAATGA 20364

Qy   1610 cgatcttggcgacatgagcggtcaaagacgcattctctcttacgacgtgga--------- 1660
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20365 CGATCTTGGCGACATGAGCGGTCAAAGACGCATTCTCTCTTACGACGTGGAAGGTAGCTA 20424

Qy   1661 ----------------------------------------------------aggtgaagg 1669
                                                               |||||||||
Db  20425 GCCTAGCATCCAGATGTTGTTAAATATTGCTAATGGGTTTCTCCTTGATTTAGGTGAAGG 20484

Qy   1670 ttttgctgtctggcagcagagttcctggatatggctattgaacttcacactgatgcttgg 1729
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20485 TTTTGCTGTCTGGCAGCAGAGTTCCTGGATATGGCTATTGAACTTCACACTGATGCTTGG 20544

Qy   1730 atgcttggtgaaattgctggtttacggtgatccgcagctggacgcgcaaacggacgccta 1789
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20545 ATGCTTGGTGAAATTGCTGGTTTACGGTGATCCGCAGCTGGACGCGCAAACGGACGCCTA 20604

Qy   1790 ctgccagcacaggcagcgggctgacttctatttagccaaggacagtcgtctcaggccta 1849
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20605 CTGCCAGCACAGGCAGCGGGCTGACTTCTATTTTAGCCAAGGACAGTCGTCTCAGGCCTA 20664

Qy   1850 cgccggttacctcaactgtctgcatatgtttggattaagtctaccggcgtcgcgcttgga 1909
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20665 CGCCGGTTACCTCAACTGTCTGCATATGTTTGGATTAAGTCTACCGGCGTCGCGCTTGGA 20724

Qy   1910 gtgttacttgcagaccacgtggcagttccttcgttttctttttccatcgcctctggctggg 1969
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20725 GTGTTACTTGCAGACCACGTGGCAGTTCCTTCGTTTTCTTTTCCATCGCCTCTGGCTGGG 20784

Qy   1970 tcgggtgctgtcacggcggtccggtgggctgtttagcaacgccgccagcaggaaacaggc 2029
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20785 TCGGGTGCTGTCACGGCGGTCCGGTGGGCTGTTTAGCAACGCCGCCAGCAGGAAACAGGC 20844

Qy   2030 gctggcatctgcacgcgaactggccctgctcttcaaccgactgaatcaattgcaactgac 2089
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20845 GCTGGCATCTGCACGCGAACTGGCCCTGCTCTTCAACCGACTGAATCAATTGCAACTGAC 20904

Qy   2090 tggaaatggaagccgcggtgacatgaacggcattatgatggcactattcgcaagcaacat 2149
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20905 TGGAAATGGAAGCCGCGGTGACATGAACGGCATTATGATGGCACTATTCGCAAGCAACAT 20964

Qy   2150 ggctgaagtggcgcacaatctactgacaccgcgcgagaccatctgcatccacgtaatgac 2209
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20965 GGCTGAAGTGGCGCACAATCTACTGACACCGCGCGAGACCATCTGCATCCACGTAATGAC 21024

Qy   2210 agcgttgcgaatgaagcgcagtgccccaaaatggttgcaacagttcttcgcccgatacta 2269
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21025 AGCGTTGCGAATGAAGCGCAGTGCCCCAAAATGGTTGCAACAGCTCTTCGCCCGATACTA 21084

Qy   2270 catgagccgggctcgtcaagagtgcggtcgcactagggccaccgagcaaacgcaggagct 2329
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21085 CATGAGCCGGGCTCGTCAAGAGTGCGGTCTCACTAGGGCCACCGAGCAAACGCAGGAGCT 21144

Qy   2330 acgttgggcattcacagcctatggatatcgctactgcgccacgcacgtcttcacgtacga 2389
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21145 ACGTTGGGCATTCACAGCCTATGGATATCGCTACTGCGCCACGCACGTCTTCACGTACGA 21204

Qy   2390 tctgagcgactccggcgagcaggatggattcttcacacgtcttaggaatccatgtgatcc 2449
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21205 TCTGAGCGACTCCGGCGAGCAGGATGGATTCTTCACACGTCTTAGGAATCCATGTGATCC 21264

Qy   2450 cgctgcccacgtcatta---------------------------------------------- 2466
          |||||||||||||||||
Db  21265 CGCTGCCCACGTCATTAAGGTAAGTTGGACTGAATTTCTAGTGTATCGCACACAATTCAA 21324

Qy   2467 --------------------------agcaatatcgagagcatttgctgtttaaatccat 2500
                                       ||||||||||||||||||||||||||| ||||
Db  21325 AACTAATTGTTTCTTTTCCAATAATCAGCAATATCGAGAGCATTTGCTGTTTAAAGCCAT 21384

Qy   2501 tcagtgtctggtaggagcgggccacaaatcgggaggcctgcccacatcttctgtcagcgg 2560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21385 TCAGTGTCTGGTAGGAGCGGGCCACAAATCGGGAGGCCTGCCCACATCTTCTGTCAGCGG 21444

Qy   2561 agaggcggaacagttgcagcaacagcagcacagcggcaccattgtcagcaatgttcttaa 2620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21445 AGAGGCGGAACAGTTGCAGCAACAGCAGCACAGCGGCACCATTGTCAGCAATGTTCTTAA 21504

Qy   2621 gtacacgtccctccttaaggacactctctgggctgatgaggatgagcgggatacaaacgt 2680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21505 GTACACGTCCCTCCTTAAGGACACTCTCTGGGCTGATGAGGATGAGCGGGATACAAACGT 21564

Qy   2681 ggtgtggtgggccgatgttttggagaccgcagtgcactggctccttggtgaagacacgct 2740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21565 GGTGTGGTGGGCCGATGTTTTGGAGACCGCAGTGCACTGGCTCCTTGGTGAAGACACGCT 21624

Qy   2741 ggccgagcaattgtacggcaggatcaagcaaatgcccacgcagctgcaacagtgcggcga 2800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21625 GGCCGAGCAATTGTACGGCAGGATCAAGCAAATGCCCACGCAGCTGCAACAGTGCGGCGA 21684

Qy   2801 aaacgatcatctgcccaaggcgctgcatgctgtgctgcgagctaagatgatcttactaaa 2860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21685 AAACGATCATCTGCCCAAGGCGCTGCATGCTGTGCTGCGAGCTAAGATGATCTTACTAAA 21744

Qy   2861 ------------------------------------------------------------ 2860

Db  21745 GTGAGTAAAACTAGAACAACTAATGAAAAGAACAAAACTAATCATCGCACACCTTTGCAG 21804

Qy   2861 aaacaatggcaacgcactggacaaaagtctcaagcaattggtaaacatcctctgcgatga 2920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21805 AAACAATGGCAACGCACTGGACAAAAGTCTCAAGCAATTGGTAAACATCCTCTGCGATGA 21864

Qy   2921 gtcgagtgtggagctccaagagtgcttgactgtcaaccggatcaccgacgccaagggtat 2980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21865 GTCGAGTGTGGAGCTCCAAGAGTGCTTGACTGTCAACCGGATCACCGACGCCAAGGGTAT 21924

Qy   2981 aaagct---------------------------------------------------- 2986
          ||||||
Db  21925 AAAGCTGGTAAGAATGTGTGCACCCCGAACAATCAGTTTTTGGCGACTGAGTGCCACCAG 21984
```

```
Qy   2987 ------------------------------------------------------------ 2986

Db  21985 ACTCTTCCTCTGGAGATGACACGAAGGTGCATTGTAGTCGCATTGTCTGTCCCAATTGCT 22044

Qy   2987 ------------------------------------------------------------ 2986

Db  22045 TCAGGCAATACAACTTCAGTGCAAGCTCTGTGCATTTCACATACAACGCCGAATGTCTCT 22104

Qy   2987 ----------------------------------------------------gcttttcc 2994
                                                               ||||||||
Db  22105 GGCAAAAGTCGCCAAGTTTTAAGAGCGCACTAATCATTATTTCTTCTATGCAGCTTTTCC 22164

Qy   2995 agttgcttacctgcgattggctgctcgaaactaggactgctctgtgggaactggaacaca 3054
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22165 AGTTGCTTACCTGCGATTGGCTGCTCGAAACTAGGACTGCTCTGTGGGAACTGGAACACA 22224

Qy   3055 tgaatatggaggacgatggcttctaccaagtgccaggtgaagtgctcgagaagttccaga 3114
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22225 TGAATATGGAGGACGATGGCTTCTACCAAGTGCCAGGTGAAGTGCTCGAGAAGTTCCAGA 22284

Qy   3115 ccgatttgaactcgttgcgcaacattgtggagaatatacc---------------------- 3154
          |||||||||||||||||||||||||||||||||||
Db  22285 CCGATTTGAACTCGTTGCGCAACATTGTGGAGAATATACCGGTATATGCAAATCCTTACC 22344

Qy   3155 ----------------------------------------------------gaacgcc 3161
                                                               |||||||
Db  22345 CTTCCAATATGTCTGTTCCGTAACCTGAGCTTGTGTAAACTTTTGCCATTACAGAACGCC 22404

Qy   3162 caatcgcgcatatatttgtacgaggcagtttgtcgcctgatggctggagcctcaccgtgt 3221
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22405 CAATCGCGCATATATTTGTACGAGGCAGTTTGTCGCCTGATGGCTGGAGCCTCACCGTGT 22464

Qy   3222 ccaacgcaacagctcttggacaggagtctgcgatcacgcaacgcccactcgtccatcttc 3281
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22465 CCAACGCAACAGCTCTTGGACAGGAGTCTGCGATCACGCAACGCCCACTCGTCCATCTTC 22524

Qy   3282 tgcggcagcaaggatcggcggcagcagaacttcgtgggcggagagcgggaacgggcttcg 3341
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22525 TGCGGCAGCAAGGATCGGCGGCAGCAGAACTTCGTGGGCGGAGAGCGGGAACGGGCTTCG 22584

Qy   3342 gccatgtacgtggcctgcaagtatctcccgcctgcgctgctcagctccccgggtgaacgt 3401
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22585 GCCATGTACGTGGCCTGCAAGTATCTCCCGCCTGCGCTGCTCAGCTCCCCGGGTGAACGT 22644

Qy   3402 gctggcatgttagccgaggcggccaagaccctggagaaggtgggcgacaagcgaaagctc 3461
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22645 GCTGGCATGTTAGCCGAGGCGGCCAAGACCCTGGAGAAGGTGGGCGACAAGCGAAAGCTC 22704

Qy   3462 aaggagtgctaccagctgatgaagtcgctgggcaacggcattggcagcgtgaaggcttag 3521
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22705 AAGGAGTGCTACCAGCTGATGAAGTCGCTGGGCAACGGCATTGGCAGCGTGAAGGCTTAG 22764

Qy   3522 gatagtagtgaagtacataataagtggcacgaacgtggtgtggatttttcagcaaatgaat 3581
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22765 GATAGTAGTGAAGTACATAATAAGTGGCACGAACGTGGTGTGGATTTTCAGCAAATGAAT 22824

Qy   3582 acccgtttgctattcaaaagaattacaaatgcctaggtctttataattacgctattcctc 3641
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22825 ACCCGTTTGCTATTCAAAAGAATTACAAATGCCTAGGTCTTTATAATTACGCTATTCCTC 22884

Qy   3642 tgttttccacgcccggttatgcttagattgtaattttaaaattatttaatatggacattt 3701
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22885 TGTTTTCCACGCCCGGTTATGCTTAGATTGTAATTTTAAAATTATTTAATATGGACATTT 22944

Qy   3702 tatttgtttattatttaccgtacttgttaaacgtatttataacaataaatatttttaacag 3761
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22945 TATTTGTTTATTATTTACCGTACTTGTTAAACGTATTTATAACAATAAATATTTTTAACAG 23004

Qy   3762 atttaaa 3768
          ||||| |
Db  23005 ATTTACA 23011
```

```
RESULT   3
AE003515/c
LOCUS       AE003515   273521 bp    DNA             INV       22-MAR-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386050 section 2
            of 54, complete sequence.
ACCESSION   AE003515 AE002602
VERSION     AE003515.1  GI:7293706
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 273521)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 273521)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
FEATURES             Location/Qualifiers
     source          1. .273521
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="3L"
     mRNA            join(<22612. .22692,22755. .22918,23186. .23319,23379. .23449,
                     23505. .>23597)
                     /gene="CG14184"
                     /product="CT33793"
                     /db_xref="FLYBASE:FBan0014184"
                     /db_xref="FLYBASE:FBgn0036932"
                     /evidence=not_experimental
     gene            <22612. .>23597
```

```
                /gene="CG14184"
                /db_xref="FLYBASE:FBan0014184"
                /db_xref="FLYBASE:FBgn0036932"
                /evidence=not_experimental
     CDS        join(22612. .22692,22755. .22918,23186. .23319,23379. .23449,
                23505. .23597)
                /gene="CG14184"
                /note="CG14184 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0014184"
                /db_xref="FLYBASE:FBgn0036932"
                /evidence=not_experimental
                /protein_id="AAF49076.1"
                /db_xref="GI:7293707"
                /translation="MEYVHSVWSYLASICNCWQCTEDPGAQPNEPNERTHLLVDPVNH
                SPALRRSNSDGLSTDYAHSLPKKDDQNALSRLVQNTAINMINVGAMDCHSLEHQEYAD
                RIRLYSQRLHQQWNNGQHASIAPKGLLKDVPSHQFYLSKPTYPDDTAQMKLFTEKAHI
                SVSHIQIDHKEAVVVPFRIP"
     mRNA       join(23732. .24542,24645. .24917,24975. .25264,25327. .25483,
                25532. .25795,25867. .26024,26109. .26692,26754. .>26858)
                /gene="CG14183"
                /product="CT33792"
                /db_xref="FLYBASE:FBan0014183"
                /db_xref="FLYBASE:FBgn0036931"
     gene       <23732. .>26858
                /gene="CG14183"
                /db_xref="FLYBASE:FBan0014183"
                /db_xref="FLYBASE:FBgn0036931"
     CDS        join(23803. .24542,24645. .24917,24975. .25264,25327. .25483,
                25532. .25795,25867. .26024,26109. .26692,26754. .26858)
                /gene="CG14183"
                /note="CG14183 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0014183"
                /db_xref="FLYBASE:FBgn0036931"
                /protein_id="AAF49077.1"
                /db_xref="GI:7293708"
                /translation="MSFDLNHKFWVYTRKEIGKLIKKQNRLKKIEPPEEKAVSYKIYA
                ELYVLYVELVNKLSFIYHNTFQVQKRALVRVLVENATQQLMQLKEELKDLEMSEYIYI
                DKALIARKLTPRDLVVWRSPQFLYRRPLDIQNIIADNRLYMNDEEKEEKAAQDWVPIS
                EAVKLIQAHERARRARVYKSNIKYDKKKFLKVHQRKKINYKFTFKPDQAMSIPVKRTI
                FSADFLKADESCENLIKKDDVADTAIDGNCWRGYKTRKIVKIRKRFKEELYGMKKIRK
                LKRPNQTANSIMEMYKNEMLKKKLDEDFINLINDERTRLLQFRSPWMMEDISDHIRAW
                FKEFYDATGNFHPYPDPVKQGTVLVVIDETMTPMEFQESLGKKPLSKAELKKLRDKEK
                KEKRKKKEKLKQLKMKEAKRRKKLRDAGVIDIGYELTASKAIENIEETMKKYSKDWRN
                VDEYLNKNHDPIKEWVTEEQLAKIHQEVRGLVDEYMRVEYELLREALAKDNDEKYKAL
                KVKYPKKKKEKRKKKPKDMTGDRTLESLYNELKDAGIIEEIGHRDFDEFITDFNFVAD
                DTRDEDGLTTLGPAKGDIKMVIQESMLGMGEFDVPKPKSLLLIGPLNSGKRLLCEIIA
                SELDAVFMNLSPENTYKYAKDLKYFLHVILKVAKAFQPTIMYIEEAHRLFWKKVPPEA
                VEINPRLLGTSLASKILKPLKKNDKIVLVGTSNMPWAANGGIKRAFQKVLLIPKCDYG
                TSFLLWLELMTENAPDDMKEYAYSALARVLQAYNSGDIDDNVKETLNIDRKMRLKNEA
                LDPNEFLEYFLSKNEPPIFPPDEKIMDKFNKWFGSSNKLEKLRKSYMAQKLAKASKKK
                "
     mRNA       complement(join(<26961. .27216,27925. .28113,28172. .28453,
                28516. .33820,33883. .34191,34257. .35155,35218. .35395,
                35457. .36186,36244. .36506,36563. .36737,36797. .36982,
                37038. .38046,38109. .38481,38544. .38699,38764. .38825,
                38881. .39088,39147. .39253,39312. .42729))
                /gene="CG7749"
                /product="CT23445"
                /db_xref="FLYBASE:FBan0007749"
                /db_xref="FLYBASE:FBgn0036930"
     gene       complement(<26961. .>42729)
                /gene="CG7749"
                /db_xref="FLYBASE:FBan0007749"
                /db_xref="FLYBASE:FBgn0036930"
     CDS        complement(join(27134. .27216,27925. .28113,28172. .28453,
                28516. .33820,33883. .34191,34257. .35155,35218. .35395,
                35457. .36186,36244. .36506,36563. .36737,36797. .36982,
                37038. .38046,38109. .38481,38544. .38699,38764. .38825,
                38881. .39088,39147. .39253,39312. .42729))
                /gene="CG7749"
                /note="CG7749 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0007749"
                /db_xref="FLYBASE:FBgn0036930"
                /protein_id="AAF49078.1"
                /db_xref="GI:7293709"
                /translation="MKIKKYVTPVKRKAFTILQWISLLCSLWLIPTVQSKADEKHTAT
                LEYRLENQLQDLYRFSHSVYNVTIPENSLGKTYAKGVLHERLAGLRVGLNAEVKYRII
                SGDKEKLFKAEEKLVGDFAFLAIRTRTNNVVLNREKTEEYVIRVKAHVHLHDRNVSSY
                ETEANIHIKVLDRNDLSPLFYPTQYTVVIPEDTPKYQSILKVTADDADLGINGEIYYS
                LLMDSEYFAIHPTTGEITLLQQLQYAENSHFELTVVAYDRGSWVNHQNHQASKTKVSI
                SVKQVNFYAPEIFTKTFSSVTPTSNPLIYGIVRVNDKDTGINGNIGRLEIVDGNPDGT
                FLLKAAETKDEYYIELNQFAHLNQQHFIYNLTLLAEDLGTPRRFAYKSVPIQIKPESK
                NIPIFTQEIYEVSIPETAPINMPVIRLKVSDPDLGKNALVYLEIVGGNEGDEFRINPD
                SGMLYTAKQLDAEKKSSYTLTVSAIDQANVGSRKQSSAKVKISVQDMNDNDPIFENVN
                KVISINENNLAGSFVVKLTAKDRDSGENSYISYSIANLNAVPFEIDHFSGIVKTTSLL
                DFETMKRNYELIIRASDWGLPYRRQTEIKLSIVVKDINDNRPQFERVNCYGKVTKSAP
                MGTEVFVTSAIDFDAGDIISYRLSDGNEDGCFNLDPTSGSLSISCDLKKTTLTNRILK
                VSATDGTHFSDDLIINVHLMPEDLGGDSSILHGFGSFECRETGVARRLAETLSLAEKN
                NVKSASPSVFSDLSLTPSRYGQNVHRPEFVNFPQELSINESVQLGETVAWIEAKDRDL
                GYNGKLVFAISDGDYDSVFRIDPDRGELQIIGYLDRERQNEYVLNITVYDLGNPTKST
                SKMLPITILDVNDNRPVIQKTLATFRLTESARIGTVVHCLHATDADSGINAQVTYALS
                VECSDFTVNATTGCLRLNKPLDREKQDNYALHITAKDGGSPVLSSEALVYVLVDDVND
                NAPVFGVQEYIFKVREDLPRGTVLAVIEAVDEDIGPNAEIQFSLKEETQDEELFRIDK
                HTGAIRTQGYLDYENKQVHNLIVSAIDGGDPSLTSDMSIVIMIIDVNENRFAPEFDDF
                VYEGKVKENKPKGTFVMNVTARDMDTVDLNSKITYSITGGDGLGIFAVNDQGSITSLS
                QLDAETKNFYWLTLCAQDCAIVPLSNCVEVYIQVENENDNIPLTDKPVYYVNVTEASV
                ENVEIITLKAFDPDIDPTQTITYNIVSGNLVGYFEIDSKTGVIKTTERKLDRENQAEH
                ILEVAISDNGSPVLSSTSRIVVSVLDINDNSPEFDQRVYKVQVPSSATVNQSIFQVHA
                IDSDSGENGRITYSIKSGKGKNKFRIDSQRGHIHIAKPLDSDNEFEIHIKAEDNGIPK
                KSQTARVNIVVVPVNPNSQNAPLIVRKTSENVVDLTENDKPGFLVTQILAVDDDNDQL
                WYNISNGNDDNTFYIGQDNGNILLSKYLDYETQQSYNLTISVTDGTFTAFTNLLVQVI
                DINDNPPQFAKDVYHVNISENIEEESVIMQLHATDRDEDKKLFYHLHATQDPSSLALF
                RIDSISGNVIVTQRLDFEKTAQHILIVFVKDQGAPGKRNYAKIIVNVHDHNDHHPEFT
                AKIIQSKVPESAAIGSKLAEVRAIDRDSGHNAEIQYSIITGNVGSVFEIDPTFGIITL
                AGNLNINKIQEYMLQVKAVDLGNPPLSSQIPVHIIVTMSENDPPKFPTNNIAIEIFEN
                LPIGTFVTQVTARSSSSIFFNIISGNINESFRINPSTGVIVINGNIDYESIKVFNLTV
                KGTNMAAESSCQNIIIHILDANDNIPYFVQNEYVGALPESAAIGSYVLKVHDSSKDHL
                TLQVKDADVGVNGMVEYHIVDDLAKNFFKIDSTTGAIELLRQLDYETNAGYTFDVTVS
                DMGKPKLHSTTTAHVTIRVINVNDCPPVFNERELNVTLFLPTFENVFVRQVSAKDADN
                DTLRFDIVDGNTNECFQIEKYTGIITTRNFEILNNENDRDYALHVRASDGIFSAILIV
                KIKVLSAIDSNFAFQRESYRFSAFENNTKVATIGLVNVIGNTLDENVEYRILNPTQLF
                DIGISSGALKTTGVIFDREVKDLYRLFVEAKSMLYDGMNSNVRRAVTSIDISVLDVND
                NCPLFVNMPYYATVSIDDPKGTIIMQVKAIDLDSAENGEVRYELKKGNGELFKLDRKS
                GELSIKQHVEGHNRNYELTVAAYDGAITPCSSEAPLQVKVIDRSMPVFEKQFYTVSVK
                EDVEMYSALSVSIEAESPLGRSLIYTISSESQSFEIDYNTGSIFVVNELDYEKISSHD
                VSIRATDSLSGVYAEVVLSVSIMDVNDCYPEIESDIYNLTIPENASFGTQILKINATD
                NDSGANAKLSYYIESINGQNNSELFYIDVTDGNLYLKTPLDYEQIKYHHIVVNVKDHG
                SPSLSSRSNVFITGRILCRFISYKLIYDSIIPVKDLNDNAPCFVEPSYFTKVSVAAVR
                GQFVALPKAYDKDISDTDSLEYKIVYGNELQTYSIDKLTGVISLQNMLNFTDKSSTVL
                NISVSDGVHTAYARLKISLLPENVYSPLFDQSTYEAQVPENLLHGHNIITVKASDGDF
                GTYANLYYEIVSEEMKKIFLIDQTTGVITSKVTFDREKKDEYVVLLKVSDGGGKFGFA
                SLKVIVVDVNDNVPYFLLKEYKMVVSTTVEANQTILTVKAKDDDIVDNGSVHFQIVQK
                SNDKAVKDVIEINEKTGDIVFKSKAESYGVNSYQFFVRASDRGEPQFHSEVPVSIEII
                ETDANIPTFEKSSVLLKIIESTPPGTVLTKLHMIGNYTFKFSIAADQDHFMISDSGEL
                ILQQTLDREQQESHNLIVVAETSTVPVFFAYADVLIDVRDENDNYPKFDNTFYSASVA
                ENSEKVISLVKVSATDADTGPNGDIRYYLESDTENIQNIFDIDIYSGWITLLTSLDRE
```

```
  Query Match             81.0%;  Score 3053;  DB 34;  Length 273521;
  Best Local Similarity   87.9%;  Pred. No. 0;
  Matches 3594;  Conservative    0;  Mismatches    5;  Indels  488;  Gaps    5;

Qy     170 agcattcgcaatggacacgacactgatgaacttaatagacgctccgctggacgagtccat 229
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  256390 AGCATTCGCAATGGACACGACACTGATGAACTTAATAGACGCTCCGCTGGACGAGTCCAT 256331

Qy     230 ggatttgttcaaagcggaggatgtcttcgaaccgttcgacgccgacctgcactcggacat 289
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  256330 GGATTTGTTCAAAGCGGAGGATGTCTTCGAACCGTTCGACGCCGACCTGCACTCGGACAT 256271

Qy     290 gctggacatcatcctcaacgacatggacctggcgccgacgcagatgtacaacatgctgct 349
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  256270 GCTGGACATCATCCTCAACGACATGGACCTGGCGCCGACGCAGATGTACAACATGCTGCT 256211
```

```
Qy    350 ggacgagcctcgaacgcatacccagcagacgcagtccgtggatcagcagccgcaatccgt 409
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 256210 GGACGAGCCTCGAACGCATACCCAGCAGACGCAGTCCGTGGATCAGCAGCCGCAATCCGT 256151

Qy    410 cgagcaacagccgcacgtgaaaagcgagcactcttcgccagtgcacatcaaggaggaact 469
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 256150 CGAGCAACAGCCGCACGTGAAAAGCGAGCACTCTTCGCCAGTGCACATCAAGGAGGAACT 256091

Qy    470 gcatcagcagcaacaacagtcgccgcttctcgtctacaaaccagatcccctcatagccac 529
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 256090 GCATCAGCAGCAACAACAGTCGCCGCTTCTCGTCTACAAACCAGATCCCCTCATAGCCAC 256031

Qy    530 aagctacaattgtccccagcaacagccgacgggccttttgaaggccgcccaaccaacagc 589
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 256030 AAGCTACAATTGTCCCCAGCAACAGCCGACGGGCCTTTTGAAGGCCGCCCAACCAACAGC 255971

Qy    590 caccatacatcacatggacgcccagcggatgccgccgaacacggcggtgtatccccccatc 649
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 255970 CACCATACATCACATGGACGCCCAGCGGATGCCGCCGAACACGGCGGTGTATCCCCCATC 255911

Qy    650 tctgggcagtagctttgtctaccagtccatgtccccgcccacgtcgccggtggagtctgc 709
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 255910 TCTGGGCAGTAGCTTTGTCTACCAGTCCATGTCCCCGCCCACGTCGCCGGTGGAGTCTGC 255851

Qy    710 gaaccagaatgtcaatgtcatgcagcccgttgctgcaactcctgctcccgcttctgctcc 769
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 255850 GAACCAGAATGTCAATGTCATGCAGCCCGTTGCTGCAACTCCTGCTCCCGCTTCTGCTCC 255791

Qy    770 tttgccccagcagtcgtatccgcaacccttcattacgtacaactctaaggccggaatgac 829
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 255790 TTTGCCCCAGCAGTCGTATCCGCAACCCTTCATTACGTACAACTCTAAGGCCGGAATGAC 255731

Qy    830 ttccgatgaagccatgtacttgctcttgcagcccacggtagccagtccaaccccatctcc 889
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db 255730 TTCCGATGAAGCCATGTACTTCCTCTTGCAGCCCACGGTAGCCAGTCCAACCCCATCTCC 255671

Qy    890 acctgtggctccaccaccgacaagcacaggtagtcgggccagcaaggtgcgagtggcacc 949
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 255670 ACCTGTGGCTCCACCACCGACAAGCACAGGTAGTCGGGCCAGCAAGGTGCGAGTGGCACC 255611

Qy    950 actggctccgtcacctgccgctatggaagtccagggcaaggtacctatcaaccgggttca 1009
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 255610 ACTGGCTCCGTCACCTGCCGCTATGGAAGTCCAGGGCAAGGTACCTATCAACCGGGTTCA 255551

Qy   1010 acccaaggtgaaggaagtaaagcgctcggcccacaacgccatcgagcggcgctatcgcac 1069
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 255550 ACCCAAGGTGAAGGAAGTAAAGCGCTCGGCCCACAACGCCATCGAGCGGCGCTATCGCAC 255491

Qy   1070 ctcaatcaacgacaagattaacgagttgaagaacttggtagtgggagagcaggccaagct 1129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 255490 CTCAATCAACGACAAGATTAACGAGTTGAAGAACTTGGTAGTGGGAGAGCAGGCCAAGCT 255431

Qy   1130 gaacaagtccgcagtgttgcggaaatccatagacaagattcgggatctgcaacgccagaa 1189
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 255430 GAACAAGTCCGCAGTGTTGCGGAAATCCATAGACAAGATTCGGGATCTGCAACGCCAGAA 255371

Qy   1190 tcacgatctgaaggcagagttgcagcgcctgcagagggagctaatggcacgcgacggctc 1249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 255370 TCACGATCTGAAGGCAGAGTTGCAGCGCCTGCAGAGGGAGCTAATGGCACGCGACGGCTC 255311

Qy   1250 caaggtgaaggatttacttcagctgggcactcggcctggtagagcatccaagaagcgccg 1309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 255310 CAAGGTGAAGGATTTACTTCAGCTGGGCACTCGGCCTGGTAGAGCATCCAAGAAGCGCCG 255251

Qy   1310 cgagagctcgcagacctttaccacggatgccggactgacgccgccacgcagcgatgaatc 1369
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 255250 CGAGAGCTCGCAGACCTTTACCACGGATGCCGGACTGACGCCGCCACGCAGCGATGAATC 255191

Qy   1370 ggatccttcgctctcgcccatgcactcggacatctcgttgccgccatcaccctatggtgg 1429
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 255190 GGATCCTTCGCTCTCGCCCATGCACTCGGACATCTCGTTGCCGCCATCACCCTATGGTGG 255131

Qy   1430 atccaccgccagctgtagcagtggcagcagcagcaatgaagaaccactggtggtgcc 1489
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 255130 ATCCACCGCCAGCTGTAGCAGTGGCAGCAGCAGCAATGAAGAACCACTGGTGGTGCC 255071

Qy   1490 cagctctatgcgcggcatggccacccactctcgcctcggactctgcatgtttatgttcgc 1549
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 255070 CAGCTCTATGCGCGGCATGGCCACCCACTCTCGCCTCGGACTCTGCATGTTTATGTTCGC 255011

Qy   1550 catcctggccgtcaatcccttcaagacctttctccagcgcggccactatgacagtaatga 1609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 255010 CATCCTGGCCGTCAATCCCTTCAAGACCTTTCTCCAGCGCGGCCACTATGACAGTAATGA 254951

Qy   1610 cgatcttggcgacatgagcggtcaaagacgcattctctcttacgacgtgga--------- 1660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 254950 CGATCTTGGCGACATGAGCGGTCAAAGACGCATTCTCTCTTACGACGTGGAAGGTAGCTA 254891

Qy   1661 ----------------------------------------------------aggtgaagg 1669
                                                               |||||||||
Db 254890 GCCTAGCATCCAGATGTTGTTAAATATTGCTAATGGGTTTCTCCTTGATTTAGGTGAAGG 254831

Qy   1670 ttttgctgtctggcagcagagttcctggatatggctattgaacttcacactgatgcttgg 1729
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 254830 TTTTGCTGTCTGGCAGCAGAGTTCCTGGATATGGCTATTGAACTTCACACTGATGCTTGG 254771

Qy   1730 atgcttggtgaaattgctggtttacggtgatccgcagctggacgcgcaaacggacgccta 1789
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 254770 ATGCTTGGTGAAATTGCTGGTTTACGGTGATCCGCAGCTGGACGCGCAAACGGACGCCTA 254711

Qy   1790 ctgccagcacaggcagcgggctgacttctatttagccaaggacagtcgtctcaggccta 1849
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 254710 CTGCCAGCACAGGCAGCGGGCTGACTTCTATTTTAGCCAAGGACAGTCGTCTCAGGCCTA 254651

Qy   1850 cgccggttacctcaactgtctgcatatgtttggattaagtctaccggcgtcgcgcttgga 1909
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 254650 CGCCGGTTACCTCAACTGTCTGCATATGTTTGGATTAAGTCTACCGGCGTCGCGCTTGGA 254591

Qy   1910 gtgttacttgcagaccacgtggcagttccttcgttttcttttccatcgcctctggctggg 1969
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 254590 GTGTTACTTGCAGACCACGTGGCAGTTCCTTCGTTTTCTTTTCCATCGCCTCTGGCTGGG 254531

Qy   1970 tcgggtgctgtcacggcggtccggtgggctgtttagcaacgccgccagcaggaaacaggc 2029
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 254530 TCGGGTGCTGTCACGGCGGTCCGGTGGGCTGTTTAGCAACGCCGCCAGCAGGAAACAGGC 254471

Qy   2030 gctggcatctgcacgcgaactggccctgctcttcaaccgactgaatcaattgcaactgac 2089
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 254470 GCTGGCATCTGCACGCGAACTGGCCCTGCTCTTCAACCGACTGAATCAATTGCAACTGAC 254411

Qy   2090 tggaaatggaagccgcggtgacatgaacggcattatgatggcactattcgcaagcaacat 2149
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 254410 TGGAAATGGAAGCCGCGGTGACATGAACGGCATTATGATGGCACTATTCGCAAGCAACAT 254351

Qy   2150 ggctgaagtggcgcacaatctactgacaccgcgcgagaccatctgcatccacgtaatgac 2209
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 254350 GGCTGAAGTGGCGCACAATCTACTGACACCGCGCGAGACCATCTGCATCCACGTAATGAC 254291

Qy   2210 agcgttgcgaatgaagcgcagtgccccaaaatggttgcaacagttcttcgcccgatacta 2269
          ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db 254290 AGCGTTGCGAATGAAGCGCAGTGCCCCAAAATGGTTGCAACAGCTCTTCGCCCGATACTA 254231

Qy   2270 catgagccgggctcgtcaagagtgcggtcgcactagggccaccgagcaaacgcaggagct 2329
          ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db 254230 CATGAGCCGGGCTCGTCAAGAGTGCGGTCTCACTAGGGCCACCGAGCAAACGCAGGAGCT 254171

Qy   2330 acgttgggcattcacagcctatggatatcgctactgcgccacgcacgtcttcacgtacga 2389
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 254170 ACGTTGGGCATTCACAGCCTATGGATATCGCTACTGCGCCACGCACGTCTTCACGTACGA 254111

Qy   2390 tctgagcgactccggcgagcaggatggattcttcacacgtcttaggaatccatgtgatcc 2449
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 254110 TCTGAGCGACTCCGGCGAGCAGGATGGATTCTTCACACGTCTTAGGAATCCATGTGATCC 254051

Qy   2450 cgctgcccacgtcatta---------------------------------------- 2466
          |||||||||||||||||
```

```
Db 254050 CGCTGCCCACGTCATTAAGGTAAGTTGGACTGAATTTCTAGTGTATCGCACACAATTCAA 253991

Qy   2467 -------------------------agcaatatcgagagcatttgctgtttaaatccat 2500
                                    |||||||||||||||||||||||||||| |||||
Db 253990 AACTAATTGTTTCTTTTCCAATAATCAGCAATATCGAGAGCATTTGCTGTTTAAAGCCAT 253931

Qy   2501 tcagtgtctggtaggagcgggccacaaatcgggaggcctgcccacatcttctgtcagcgg 2560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 253930 TCAGTGTCTGGTAGGAGCGGGCCACAAATCGGGAGGCCTGCCCACATCTTCTGTCAGCGG 253871

Qy   2561 agaggcggaacagttgcagcaacagcagcacagcggcaccattgtcaccaatgttcttaa 2620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 253870 AGAGGCGGAACAGTTGCAGCAACAGCAGCACAGCGGCACCATTGTCAGCAATGTTCTTAA 253811

Qy   2621 gtacacgtccctccttaaggacactctctgggctgatgaggatgagcgggatacaaacgt 2680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 253810 GTACACGTCCCTCCTTAAGGACACTCTCTGGGCTGATGAGGATGAGCGGGATACAAACGT 253751

Qy   2681 ggtgtggtgggccgatgttttggagaccgcagtgcactggctccttggtgaagacacgct 2740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 253750 GGTGTGGTGGGCCGATGTTTTGGAGACCGCAGTGCACTGGCTCCTTGGTGAAGACACGCT 253691

Qy   2741 ggccgagcaattgtacggcaggatcaagcaaatgcccacgcagctgcaacagtgcggcga 2800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 253690 GGCCGAGCAATTGTACGGCAGGATCAAGCAAATGCCCACGCAGCTGCAACAGTGCGGCGA 253631

Qy   2801 aaacgatcatctgcccaaggcgctgcatgctgtgctgcgagctaagatgatcttactaaa 2860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 253630 AAACGATCATCTGCCCAAGGCGCTGCATGCTGTGCTGCGAGCTAAGATGATCTTACTAAA 253571

Qy   2861 ------------------------------------------------------------ 2860

Db 253570 GTGAGTAAAACTAGAACAACTAATGAAAAGAACAAAACTAATCATCGCACACCTTTGCAG 253511

Qy   2861 aaacaatggcaacgcactggacaaaagtctcaagcaattggtaaacatcctctgcgatga 2920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 253510 AAACAATGGCAACGCACTGGACAAAAGTCTCAAGCAATTGGTAAACATCCTCTGCGATGA 253451

Qy   2921 gtcgagtgtggagctccaagagtgcttgactgtcaaccggatcaccgacgccaagggtat 2980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 253450 GTCGAGTGTGGAGCTCCAAGAGTGCTTGACTGTCAACCGGATCACCGACGCCAAGGGTAT 253391

Qy   2981 aaagct---------------------------------------------------- 2986
          ||||||
Db 253390 AAAGCTGGTAAGAATGTGTGCACCCCGAACAATCAGTTTTTGGCGACTGAGTGCCACCAG 253331

Qy   2987 ------------------------------------------------------------ 2986

Db 253330 ACTCTTCCTCTGGAGATGACACGAAGGTGCATTGTAGTCGCATTGTCTGTCCCAATTGCT 253271

Qy   2987 ------------------------------------------------------------ 2986

Db 253270 TCAGGCAATACAACTTCAGTGCAAGCTCTGTGCATTTCACATACAACGCCGAATGTCTCT 253211

Qy   2987 ----------------------------------------------------gcttttcc 2994
                                                             ||||||||
Db 253210 GGCAAAAGTCGCCAAGTTTTAAGAGCGCACTAATCATTATTTCTTCTATGCAGCTTTTCC 253151

Qy   2995 agttgcttacctgcgattggctgctcgaaactaggactgctctgtgggaactggaacaca 3054
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 253150 AGTTGCTTACCTGCGATTGGCTGCTCGAAACTAGGACTGCTCTGTGGGAACTGGAACACA 253091

Qy   3055 tgaatatggaggacgatggcttctaccaagtgccaggtgaagtgctcgagaagttccaga 3114
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 253090 TGAATATGGAGGACGATGGCTTCTACCAAGTGCCAGGTGAAGTGCTCGAGAAGTTCCAGA 253031

Qy   3115 ccgatttgaactcgttgcgcaacattgtggagaatatacc---------------------- 3154
          ||||||||||||||||||||||||||||||||||||||
Db 253030 CCGATTTGAACTCGTTGCGCAACATTGTGGAGAATATACCGGTATATGCAAATCCTTACC 252971

Qy   3155 ----------------------------------------------------gaacgcc 3161
                                                             |||||||
Db 252970 CTTCCAATATGTCTGTTCCGTAACCTGAGCTTGTGTAAACTTTTGCCATTACAGAACGCC 252911
```

```
Qy   3162 caatcgcgcatatatttgtacgaggcagtttgtcgcctgatggctggagcctcaccgtgt 3221
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 252910 CAATCGCGCATATATTTGTACGAGGCAGTTTGTCGCCTGATGGCTGGAGCCTCACCGTGT 252851

Qy   3222 ccaacgcaacagctcttggacaggagtctgcgatcacgcaacgcccactcgtccatcttc 3281
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 252850 CCAACGCAACAGCTCTTGGACAGGAGTCTGCGATCACGCAACGCCCACTCGTCCATCTTC 252791

Qy   3282 tgcggcagcaaggatcggcggcagcagaacttcgtgggcggagagcgggaacgggcttcg 3341
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 252790 TGCGGCAGCAAGGATCGGCGGCAGCAGAACTTCGTGGGCGGAGAGCGGGAACGGGCTTCG 252731

Qy   3342 gccatgtacgtggcctgcaagtatctcccgcctgcgctgctcagctccccgggtgaacgt 3401
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 252730 GCCATGTACGTGGCCTGCAAGTATCTCCCGCCTGCGCTGCTCAGCTCCCCGGGTGAACGT 252671

Qy   3402 gctggcatgttagccgaggcggccaagaccctggagaaggtgggcgacaagcgaaagctc 3461
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 252670 GCTGGCATGTTAGCCGAGGCGGCCAAGACCCTGGAGAAGGTGGGCGACAAGCGAAAGCTC 252611

Qy   3462 aaggagtgctaccagctgatgaagtcgctgggcaacggcattggcagcgtgaaggcttag 3521
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 252610 AAGGAGTGCTACCAGCTGATGAAGTCGCTGGGCAACGGCATTGGCAGCGTGAAGGCTTAG 252551

Qy   3522 gatagtagtgaagtacataataagtggcacgaacgtggtgtggattttcagcaaatgaat 3581
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 252550 GATAGTAGTGAAGTACATAATAAGTGGCACGAACGTGGTGTGGATTTTCAGCAAATGAAT 252491

Qy   3582 acccgtttgctattcaaaagaattacaaatgcctaggtctttataattacgctattcctc 3641
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 252490 ACCCGTTTGCTATTCAAAAGAATTACAAATGCCTAGGTCTTTATAATTACGCTATTCCTC 252431

Qy   3642 tgttttccacgcccggttatgcttagattgtaattttaaaattatttaatatggacattt 3701
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 252430 TGTTTTCCACGCCCGGTTATGCTTAGATTGTAATTTTAAAATTATTTAATATGGACATTT 252371

Qy   3702 tatttgtttattatttaccgtacttgttaaacgtatttataacaataaatattttaacag 3761
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 252370 TATTTGTTTATTATTTACCGTACTTGTTAAACGTATTTATAACAATAAATATTTTAACAG 252311

Qy   3762 atttaaa 3768
          ||||| |
Db 252310 ATTTACA 252304
```

```
RESULT   4
AC010021/c
LOCUS       AC010021    57873 bp    DNA             HTG       18-FEB-2000
DEFINITION  Drosophila melanogaster clone RPCI98-11A8, *** SEQUENCING IN
            PROGRESS ***, 46 unordered pieces.
ACCESSION   AC010021
VERSION     AC010021.3  GI:6996706
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 57873)
  AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
            Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
            Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
            Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
            Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
```

```
            Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
            Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
            Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
            Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,
            Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 57873)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Feb 18, 2000 this sequence version replaced gi:5881473.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: DRBU
                Center clone name: RPCI98-11A8
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 99% of reads
                Assembly program: Phrap; version 0.980611
                Consensus quality: 25215 bases at least Q40
                Consensus quality: 39511 bases at least Q30
                Consensus quality: 46436 bases at least Q20
                Estimated insert size: 45078; sum-of-contigs estimation
                Quality coverage: 0.7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 46 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      831: contig of 831 bp in length
            *      832      851: gap of unknown length
            *      852     1687: contig of 836 bp in length
            *     1688     1707: gap of unknown length
            *     1708     2546: contig of 839 bp in length
            *     2547     2566: gap of unknown length
            *     2567     3384: contig of 818 bp in length
            *     3385     3404: gap of unknown length
            *     3405     4818: contig of 1414 bp in length
            *     4819     4838: gap of unknown length
            *     4839     5673: contig of 835 bp in length
            *     5674     5693: gap of unknown length
            *     5694     6598: contig of 905 bp in length
            *     6599     6618: gap of unknown length
            *     6619     7540: contig of 922 bp in length
            *     7541     7560: gap of unknown length
            *     7561     8940: contig of 1380 bp in length
            *     8941     8960: gap of unknown length
            *     8961    10248: contig of 1288 bp in length
            *    10249    10268: gap of unknown length
            *    10269    11351: contig of 1083 bp in length
            *    11352    11371: gap of unknown length
            *    11372    12481: contig of 1110 bp in length
            *    12482    12501: gap of unknown length
            *    12502    14007: contig of 1506 bp in length
            *    14008    14027: gap of unknown length
            *    14028    14865: contig of 838 bp in length
            *    14866    14885: gap of unknown length
            *    14886    16199: contig of 1314 bp in length
            *    16200    16219: gap of unknown length
            *    16220    17277: contig of 1058 bp in length
            *    17278    17297: gap of unknown length
            *    17298    18246: contig of 949 bp in length
            *    18247    18266: gap of unknown length
            *    18267    19687: contig of 1421 bp in length
            *    19688    19707: gap of unknown length
            *    19708    21218: contig of 1511 bp in length
            *    21219    21238: gap of unknown length
            *    21239    22141: contig of 903 bp in length
            *    22142    22161: gap of unknown length
            *    22162    23043: contig of 882 bp in length
            *    23044    23063: gap of unknown length
            *    23064    23478: contig of 415 bp in length
            *    23479    23498: gap of unknown length
            *    23499    25055: contig of 1557 bp in length
            *    25056    25075: gap of unknown length
            *    25076    26050: contig of 975 bp in length
            *    26051    26070: gap of unknown length
            *    26071    27618: contig of 1548 bp in length
            *    27619    27638: gap of unknown length
            *    27639    28649: contig of 1011 bp in length
            *    28650    28669: gap of unknown length
            *    28670    29958: contig of 1289 bp in length
            *    29959    29978: gap of unknown length
            *    29979    30893: contig of 915 bp in length
            *    30894    30913: gap of unknown length
            *    30914    32447: contig of 1534 bp in length
            *    32448    32467: gap of unknown length
            *    32468    33518: contig of 1051 bp in length
            *    33519    33538: gap of unknown length
            *    33539    34583: contig of 1045 bp in length
            *    34584    34603: gap of unknown length
            *    34604    35754: contig of 1151 bp in length
            *    35755    35774: gap of unknown length
            *    35775    37161: contig of 1387 bp in length
            *    37162    37181: gap of unknown length
            *    37182    38633: contig of 1452 bp in length
            *    38634    38653: gap of unknown length
            *    38654    39940: contig of 1287 bp in length
            *    39941    39960: gap of unknown length
            *    39961    41986: contig of 2026 bp in length
            *    41987    42006: gap of unknown length
            *    42007    43675: contig of 1669 bp in length
            *    43676    43695: gap of unknown length
            *    43696    44832: contig of 1137 bp in length
            *    44833    44852: gap of unknown length
            *    44853    46381: contig of 1529 bp in length
            *    46382    46401: gap of unknown length
            *    46402    47470: contig of 1069 bp in length
            *    47471    47490: gap of unknown length
            *    47491    49244: contig of 1754 bp in length
            *    49245    49264: gap of unknown length
            *    49265    50791: contig of 1527 bp in length
            *    50792    50811: gap of unknown length
            *    50812    52579: contig of 1768 bp in length
            *    52580    52599: gap of unknown length
            *    52600    53959: contig of 1360 bp in length
            *    53960    53979: gap of unknown length
            *    53980    56252: contig of 2273 bp in length
            *    56253    56272: gap of unknown length
            *    56273    57873: contig of 1601 bp in length.
FEATURES             Location/Qualifiers
     source          1. .57873
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="RPCI98-11A8"
BASE COUNT    15592 a  12473 c  12485 g  16366 t    957 others
ORIGIN


  Query Match             22.6%;  Score 853.2;  DB 51;  Length 57873;
  Best Local Similarity   89.8%;  Pred. No. 1.1e-192;
  Matches  965;  Conservative    0;  Mismatches   39;  Indels   71;  Gaps    3;

Qy   1858 acctcaactgtctgcatatgtttggattaagtctaccggcgtcgcgcttggagtgttact 1917
```

```
Db  47494 ACCTNNNNNNNNNNNNNNNNNNNNNGATTAAGTCTACCGGCGTCGCGCTTGGAGTGTTACT 47435

Qy   1918 tgcagaccacgtggcagttccttcgttttcttttccatcgcctctggctgggtcgggtgc 1977
          ||||||||||||    ||||||||||||||| |     |  ||||||||||||||||
Db  47434 TGCAGACCACGTGCAGTTTCCTTCGTTTTCTTTCCATCGCCTCCTGGCTGGGTCGGGTGC 47375

Qy   1978 tgtcacggcggtccggtgggctgtttagcaacgccgccagcaggaaacaggcgctggcat 2037
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db  47374 TGTCACGGCGGTCCGGTGNGCTGTTTAGCAACGCCGCCAGCAGGAAACAGGCGCTGGCAT 47315

Qy   2038 ctgcacgcgaactggccctgctcttcaaccgactgaatcaattgcaactgactggaaatg 2097
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  47314 CTGCACGCGAACTGGCCCTGCTCTTCAACCGACTGAATCAATTGCAACTGACTGGAAATG 47255

Qy   2098 gaagccgcggtgacatgaacggcattatgatggcactattcgcaagcaacatggctgaag 2157
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  47254 GAAGCCGCGGTGACATGAACGGCATTATGATGGCACTATTCGCAAGCAACATGGCTGAAG 47195

Qy   2158 tggcgcacaatctactgacaccgcgcgagaccatctgcatccacgtaatgacagcgttgc 2217
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  47194 TGGCGCACAATCTACTGACACCGCGCGAGACCATCTGCATCCACGTAATGACAGCGTTGC 47135

Qy   2218 gaatgaagcgcagtgccccaaaatggttgcaacagttcttcgcccgatactacatgagcc 2277
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db  47134 GAATGAAGCGCAGTGCCCCAAAATGGTTGCAACAGCTCTTCGCCCGATACTACATGAGCC 47075

Qy   2278 gggctcgtcaagagtgcggtcgcactagggccaccgagcaaacgcaggagctacgttggg 2337
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db  47074 GGGCTCGTCAAGAGTGCGGTCTCACTAGGGCCACCGAGCAAACGCAGGAGCTACGTTGGG 47015

Qy   2338 cattcacagcctatggatatcgctactgcgccacgcacgtcttcacgtacgatctgagcg 2397
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  47014 CATTCACAGCCTATGGATATCGCTACTGCGCCACGCACGTCTTCACGTACGATCTGAGCG 46955

Qy   2398 actccggcgagcaggatggattcttcacacgtcttaggaatccatgtgatcccgctgccc 2457
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  46954 ACTCCGGCGAGCAGGATGGATTCTTCACACGTCTTAGGAATCCATGTGATCCCGCTGCCC 46895

Qy   2458 acgtcatta---------------------------------------------------- 2466
          |||||||||
Db  46894 ACGTCATTAAGGTAAGTTGGACTGAATTTCTAGTGTATCGCACACAATTCAAAACTAATT 46835

Qy   2467 ------------------agcaatatcgagagcatttgctgtttaaatccattcagtgtc 2508
                           |||||||||||||||||||||||||| ||||||||||||
Db  46834 GTTTCTTTTCCAATAATCAGCAATATCGAGAGCATTTGCTGTTTAAAGCCATTCAGTGTC 46775

Qy   2509 tggtaggagcgggccacaaatcgggaggcctgcccacatcttctgtcagcggagaggcgg 2568
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  46774 TGGTAGGAGCGGGCCACAAATCGGGAGGCCTGCCCACATCTTCTGTCAGCGGAGAGGCGG 46715

Qy   2569 aacagttgcagcaacagcagcacagcggcaccattgtcagcaatgttcttaagtacacgt 2628
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  46714 AACAGTTGCAGCAACAGCAGCACAGCGGCACCATTGTCAGCAATGTTCTTAAGTACACGT 46655

Qy   2629 ccctccttaaggacactctctgggctgatgaggatgagcgggatacaaacgtggtgtggt 2688
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  46654 CCCTCCTTAAGGACACTCTCTGGGCTGATGAGGATGAGCGGGATACAAACGTGGTGTGGT 46595

Qy   2689 gggccgatgttttggagaccgcagtgcactggctccttggtgaagacacgctggccgagc 2748
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  46594 GGGCCGATGTTTTGGAGACCGCAGTGCACTGGCTCCTTGGTGAAGACACGCTGGCCGAGC 46535

Qy   2749 aattgtacggcaggatcaagcaaatgcccacgcagctgcaacagtgcggcgaaaacgatc 2808
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||
Db  46534 AATTGTACGGCAGGATCAAGCAAATGCCCACGCAGCTGCAACAGTGCGGCGAAAACCTTC 46475

Qy   2809 atctgcccaaggc-gctgcatgctgtgctgcgagct-aagatgatcttactaaaa 2861
          |||||||||||| |||||||||||||||||||||| |||||||||||||| ||
Db  46474 ATCTGCCCAAGGCAGCTGCATGCTGTGCTGCGAGCTCAAGATGATCTTACTACAA 46420

RESULT  5
```

```
AC009382/c
LOCUS       AC009382    92677 bp    DNA             HTG       18-FEB-2000
DEFINITION  Drosophila melanogaster clone RPCI98-1B21, *** SEQUENCING IN
            PROGRESS ***, 52 unordered pieces.
ACCESSION   AC009382
VERSION     AC009382.4  GI:6996682
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 92677)
  AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
            Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
            Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
            Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
            Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
            Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
            Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
            Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
            Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,
            Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 92677)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Feb 18, 2000 this sequence version replaced gi:5881494.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: DRAU
                Center clone name: RPCI98-1B21
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 42% of reads
                Chemistry: Dye-terminator Big Dye: 24% of reads
                Assembly program: Phrap; version 0.980611
                Consensus quality: 38748 bases at least Q40
                Consensus quality: 54037 bases at least Q30
                Consensus quality: 64416 bases at least Q20
                Estimated insert size: 135000; agarose-fp estimation
                Estimated insert size: 83602; sum-of-contigs estimation
                Quality coverage: 0.5x in Q20 bases; agarose-fp estimation
                Quality coverage: 0.7x in Q20 bases; sum-of-contigs estimation
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 52 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      828: contig of 828 bp in length
            *      829      848: gap of unknown length
            *      849     2184: contig of 1336 bp in length
```

```
*     2185     2204: gap of unknown length
*     2205     3611: contig of 1407 bp in length
*     3612     3631: gap of unknown length
*     3632     4316: contig of 685 bp in length
*     4317     4336: gap of unknown length
*     4337     5322: contig of 986 bp in length
*     5323     5342: gap of unknown length
*     5343     6799: contig of 1457 bp in length
*     6800     6819: gap of unknown length
*     6820     7605: contig of 786 bp in length
*     7606     7625: gap of unknown length
*     7626     8704: contig of 1079 bp in length
*     8705     8724: gap of unknown length
*     8725     9498: contig of 774 bp in length
*     9499     9518: gap of unknown length
*     9519    10995: contig of 1477 bp in length
*    10996    11015: gap of unknown length
*    11016    11827: contig of 812 bp in length
*    11828    11847: gap of unknown length
*    11848    12755: contig of 908 bp in length
*    12756    12775: gap of unknown length
*    12776    13506: contig of 731 bp in length
*    13507    13526: gap of unknown length
*    13527    14778: contig of 1252 bp in length
*    14779    14798: gap of unknown length
*    14799    15619: contig of 821 bp in length
*    15620    15639: gap of unknown length
*    15640    17154: contig of 1515 bp in length
*    17155    17174: gap of unknown length
*    17175    18583: contig of 1409 bp in length
*    18584    18603: gap of unknown length
*    18604    19693: contig of 1090 bp in length
*    19694    19713: gap of unknown length
*    19714    21010: contig of 1297 bp in length
*    21011    21030: gap of unknown length
*    21031    22445: contig of 1415 bp in length
*    22446    22465: gap of unknown length
*    22466    23147: contig of 682 bp in length
*    23148    23167: gap of unknown length
*    23168    24468: contig of 1301 bp in length
*    24469    24488: gap of unknown length
*    24489    25609: contig of 1121 bp in length
*    25610    25629: gap of unknown length
*    25630    27332: contig of 1703 bp in length
*    27333    27352: gap of unknown length
*    27353    28358: contig of 1006 bp in length
*    28359    28378: gap of unknown length
*    28379    29948: contig of 1570 bp in length
*    29949    29968: gap of unknown length
*    29969    31673: contig of 1705 bp in length
*    31674    31693: gap of unknown length
*    31694    33056: contig of 1363 bp in length
*    33057    33076: gap of unknown length
*    33077    34658: contig of 1582 bp in length
*    34659    34678: gap of unknown length
*    34679    36591: contig of 1913 bp in length
*    36592    36611: gap of unknown length
*    36612    37617: contig of 1006 bp in length
*    37618    37637: gap of unknown length
*    37638    38903: contig of 1266 bp in length
*    38904    38923: gap of unknown length
*    38924    41024: contig of 2101 bp in length
*    41025    41044: gap of unknown length
*    41045    42943: contig of 1899 bp in length
*    42944    42963: gap of unknown length
*    42964    44680: contig of 1717 bp in length
*    44681    44700: gap of unknown length
*    44701    46683: contig of 1983 bp in length
*    46684    46703: gap of unknown length
*    46704    49058: contig of 2355 bp in length
*    49059    49078: gap of unknown length
*    49079    50931: contig of 1853 bp in length
*    50932    50951: gap of unknown length
*    50952    52258: contig of 1307 bp in length
*    52259    52278: gap of unknown length
*    52279    54271: contig of 1993 bp in length
*    54272    54291: gap of unknown length
*    54292    56574: contig of 2283 bp in length
*    56575    56594: gap of unknown length
*    56595    58695: contig of 2101 bp in length
*    58696    58715: gap of unknown length
*    58716    61026: contig of 2311 bp in length
*    61027    61046: gap of unknown length
*    61047    63594: contig of 2548 bp in length
*    63595    63614: gap of unknown length
*    63615    65977: contig of 2363 bp in length
*    65978    65997: gap of unknown length
*    65998    68802: contig of 2805 bp in length
*    68803    68822: gap of unknown length
*    68823    72229: contig of 3407 bp in length
*    72230    72249: gap of unknown length
*    72250    75163: contig of 2914 bp in length
*    75164    75183: gap of unknown length
*    75184    78054: contig of 2871 bp in length
*    78055    78074: gap of unknown length
*    78075    83202: contig of 5128 bp in length
*    83203    83222: gap of unknown length
*    83223    86848: contig of 3626 bp in length
*    86849    86868: gap of unknown length
*    86869    92677: contig of 5809 bp in length.
FEATURES             Location/Qualifiers
     source          1. .92677
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="RPCI98-1B21"
BASE COUNT    26406 a  19513 c  19116 g  26545 t   1097 others
ORIGIN


  Query Match           19.0%;  Score 714.6;  DB 51;  Length 92677;
  Best Local Similarity  95.4%;  Pred. No. 1.4e-159;
  Matches  745;  Conservative    0;  Mismatches   34;  Indels    2;  Gaps    2;

Qy     768 cctttgccccagcagtcgtatccgcaaccccttcattacgtacaactctaaggccggaatg 827
           |||                  | |    ||||| |||||||||||||||| ||||||
Db    9521 CCTNNNNNNNNNNNNNNNNNNGGAGTGGTTCATCACGTACAACTCTAAGGCAGGAATG 9462

Qy     828 acttccgatgaagccatgtacttgctcttgcagcccacggtagccagtccaaccccatct 887
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    9461 ACTTCCGATGAAGCCATGTACTTCCTCTTGCAGCCCACGGTAGCCAGTCCAACCCCATCT 9402

Qy     888 ccacct-gtggctccaccaccgacaagcacaggtagtcgggccagcaaggtgcgagtggc 946
           |||||| |||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    9401 CCACCTGGTGGCTCCACCACCGACAAGCACAGGTAGTCGGGCCAGCGAGGTGCGAGTGGC 9342

Qy     947 accactggctccgtcacctgccgctatggaagtccagggcaaggtacctatcaaccgggt 1006
           | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9341 AGC-CTGGCTCCGTCACCTGCCGCTATGGAAGTCCAGGGCAAGGTACCTATCAACCGGGT 9283

Qy    1007 tcaacccaaggtgaaggaagtaaagcgctcggcccacaacgccatcgagcggcgctatcg 1066
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9282 TCAACCCAAGGTGAAGGAAGTAAAGCGCTCGGCCCACAACGCCATCGAGCGGCGCTATCG 9223

Qy    1067 cacctcaatcaacgacaagattaacgagttgaagaacttggtagtgggagagcaggccaa 1126
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9222 CACCTCAGTCAACGACAAGATTAACGAGTTGAAGAACTTGGTAGTGGGAGAGCAGGCCAA 9163

Qy    1127 gctgaacaagtccgcagtgttgcggaaatccatagacaagattcgggatctgcaacgcca 1186
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9162 GCTGAACAAGTCCGCAGTGTTGCGGAAATCCATAGACAAGATTCGGGATCTGCAACGCCA 9103

Qy    1187 gaatcacgatctgaaggcagagttgcagcgcctgcagagggagctaatggcacgcgacgg 1246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9102 GAATCACGATCTGAAGGCAGAGTTGCAGCGCCTGCAGAGGGAGCTAATGGCACGCGACGG 9043

Qy    1247 ctccaaggtgaaggatttacttcagctgggcactcggcctggtagagcatccaagaagcg 1306
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    9042 CTCCAAGGTGAAGGATTTACTTCAGCTGGGCACTCGGCCTGGTACAGCATCCAAGAAGCG 8983

Qy    1307 ccgcgagagctcgcagacctttaccacggatgccggactgacgccgccacgcagcgatga 1366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8982 CCGCGAGAGCTCGCAGACCTTTACCACGGATGCCGGACTGACGCCGCCACGCAGCGATGA 8923

Qy    1367 atcggatccttcgctctcgcccatgcactcggacatctcgttgccgccatcaccctatgg 1426
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8922 ATCGGATCCTTCGCTCTCGCCCATGCACTCGGACATCTCGTTGCCGCCATCACCCTATGG 8863

Qy    1427 tggatccaccgccagctgtagcagtggcagcagcagcagcaatgaagaaccactggtggt 1486
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8862 TGGATCCACCGCCAGCTGTAGCAGTGGCAGCAGCAGCAGCAATGAAGAACCACTGGTGGA 8803

Qy    1487 gcccagctctatgcgcggcatggccacccactctcgcctcggactctgcatgtttatgtt 1546
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    8802 GCCCAGCTCTATGCGCGGCATGGCCACCCACTCTCGCCTCGGACTCTACATGTTTATGTT 8743

Qy    1547 c 1547
          |
Db    8742 C 8742
```

```
RESULT   6
AC010021
LOCUS       AC010021    57873 bp    DNA             HTG       18-FEB-2000
DEFINITION  Drosophila melanogaster clone RPCI98-11A8, *** SEQUENCING IN
            PROGRESS ***, 46 unordered pieces.
ACCESSION   AC010021
VERSION     AC010021.3  GI:6996706
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 57873)
  AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
            Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
            Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
            Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
            Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
            Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
            Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
            Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
            Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,
            Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 57873)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Feb 18, 2000 this sequence version replaced gi:5881473.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: DRBU
                Center clone name: RPCI98-11A8
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 99% of reads
                Assembly program: Phrap; version 0.980611
                Consensus quality: 25215 bases at least Q40
                Consensus quality: 39511 bases at least Q30
                Consensus quality: 46436 bases at least Q20
                Estimated insert size: 45078; sum-of-contigs estimation
                Quality coverage: 0.7x in Q20 bases; sum-of-contigs estimation
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 46 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      831: contig of 831 bp in length
            *      832      851: gap of unknown length
            *      852     1687: contig of 836 bp in length
            *     1688     1707: gap of unknown length
            *     1708     2546: contig of 839 bp in length
            *     2547     2566: gap of unknown length
            *     2567     3384: contig of 818 bp in length
            *     3385     3404: gap of unknown length
            *     3405     4818: contig of 1414 bp in length
            *     4819     4838: gap of unknown length
            *     4839     5673: contig of 835 bp in length
            *     5674     5693: gap of unknown length
            *     5694     6598: contig of 905 bp in length
            *     6599     6618: gap of unknown length
            *     6619     7540: contig of 922 bp in length
            *     7541     7560: gap of unknown length
            *     7561     8940: contig of 1380 bp in length
            *     8941     8960: gap of unknown length
            *     8961    10248: contig of 1288 bp in length
            *    10249    10268: gap of unknown length
            *    10269    11351: contig of 1083 bp in length
            *    11352    11371: gap of unknown length
            *    11372    12481: contig of 1110 bp in length
            *    12482    12501: gap of unknown length
            *    12502    14007: contig of 1506 bp in length
            *    14008    14027: gap of unknown length
            *    14028    14865: contig of 838 bp in length
            *    14866    14885: gap of unknown length
            *    14886    16199: contig of 1314 bp in length
            *    16200    16219: gap of unknown length
            *    16220    17277: contig of 1058 bp in length
            *    17278    17297: gap of unknown length
            *    17298    18246: contig of 949 bp in length
            *    18247    18266: gap of unknown length
            *    18267    19687: contig of 1421 bp in length
            *    19688    19707: gap of unknown length
            *    19708    21218: contig of 1511 bp in length
            *    21219    21238: gap of unknown length
            *    21239    22141: contig of 903 bp in length
            *    22142    22161: gap of unknown length
            *    22162    23043: contig of 882 bp in length
            *    23044    23063: gap of unknown length
            *    23064    23478: contig of 415 bp in length
            *    23479    23498: gap of unknown length
            *    23499    25055: contig of 1557 bp in length
            *    25056    25075: gap of unknown length
            *    25076    26050: contig of 975 bp in length
            *    26051    26070: gap of unknown length
            *    26071    27618: contig of 1548 bp in length
            *    27619    27638: gap of unknown length
            *    27639    28649: contig of 1011 bp in length
            *    28650    28669: gap of unknown length
            *    28670    29958: contig of 1289 bp in length
            *    29959    29978: gap of unknown length
```

```
            *    29979    30893: contig of 915 bp in length
            *    30894    30913: gap of unknown length
            *    30914    32447: contig of 1534 bp in length
            *    32448    32467: gap of unknown length
            *    32468    33518: contig of 1051 bp in length
            *    33519    33538: gap of unknown length
            *    33539    34583: contig of 1045 bp in length
            *    34584    34603: gap of unknown length
            *    34604    35754: contig of 1151 bp in length
            *    35755    35774: gap of unknown length
            *    35775    37161: contig of 1387 bp in length
            *    37162    37181: gap of unknown length
            *    37182    38633: contig of 1452 bp in length
            *    38634    38653: gap of unknown length
            *    38654    39940: contig of 1287 bp in length
            *    39941    39960: gap of unknown length
            *    39961    41986: contig of 2026 bp in length
            *    41987    42006: gap of unknown length
            *    42007    43675: contig of 1669 bp in length
            *    43676    43695: gap of unknown length
            *    43696    44832: contig of 1137 bp in length
            *    44833    44852: gap of unknown length
            *    44853    46381: contig of 1529 bp in length
            *    46382    46401: gap of unknown length
            *    46402    47470: contig of 1069 bp in length
            *    47471    47490: gap of unknown length
            *    47491    49244: contig of 1754 bp in length
            *    49245    49264: gap of unknown length
            *    49265    50791: contig of 1527 bp in length
            *    50792    50811: gap of unknown length
            *    50812    52579: contig of 1768 bp in length
            *    52580    52599: gap of unknown length
            *    52600    53959: contig of 1360 bp in length
            *    53960    53979: gap of unknown length
            *    53980    56252: contig of 2273 bp in length
            *    56253    56272: gap of unknown length
            *    56273    57873: contig of 1601 bp in length.
FEATURES             Location/Qualifiers
     source          1. .57873
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="RPCI98-11A8"
BASE COUNT    15592 a  12473 c  12485 g  16366 t    957 others
ORIGIN


  Query Match           13.9%;  Score 524.4;  DB 51;  Length 57873;
  Best Local Similarity  97.6%;  Pred. No. 3.3e-114;
  Matches  531;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy     170 agcattcgcaatggacacgacactgatgaacttaatagacgctccgctggacgagtccat 229
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   43132 AGCATTCGCAATGGACACGACACTGATGAACTTAATAGACGCTCCGCTGGACGAGTCCAT 43191

Qy     230 ggatttgttcaaagcggaggatgtcttcgaaccgttcgacgccgacctgcactcggacat 289
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   43192 GGATTTGTTCAAAGCGGAGGATGTCTTCGAACCGTTCGACGCCGACCTGCACTCGGACAT 43251

Qy     290 gctggacatcatcctcaacgacatggacctggcgccgacgcagatgtacaacatgctgct 349
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   43252 GCTGGACATCATCCTCAACGACATGGACCTGGCGCCGACGCAGATGTACAACATGCTGCT 43311

Qy     350 ggacgagcctcgaacgcatacccagcagacgcagtccgtggatcagcagccgcaatccgt 409
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   43312 GGACGAGCCTCGAACGCATACCCAGCAGACGCAGTCCGTGGATCAGCAGCCGCAATCCGT 43371

Qy     410 cgagcaacagccgcacgtgaaaagcgagcactcttcgccagtgcacatcaaggaggaact 469
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   43372 CGAGCAACAGCCGCACGTGAAAAGCGAGCACTCTTCGCCAGTGCACATCAAGGAGGAACT 43431

Qy     470 gcatcagcagcaacaacagtcgccgcttctcgtctacaaaccagatcccctcatagccac 529
           ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   43432 GCATCAGCAGCAACAACAGTCGCCGCTTCTCGTCTACAAACCAGATTCCCTCATAGCCAC 43491

Qy     530 aagctacaattgtccccagcaacagccgacgggccttttgaaggccgcccaaccaacagc 589
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   43492 AAGCTACAATTGTCCCCAGCAACAGCCGACGGGCCTTTTGAAGGGCGCCCAACCAACAGC 43551

Qy     590 caccatacatcacatggacgcccagcggatgccgccgaacacggcggtgtatcccccatc 649
           ||||||||||||||||||||||||||||||||||||||||||||||||||||  |||||||
Db   43552 CACCATACATCACATGGACGCCCAGCGGATGCCGCCGAACACGGCGGTGTANTCCCCATC 43611

Qy     650 tctgggcagtagctttgtctaccagtccatgtccccgcccacgtcgccggtggagtctgc 709
           |||||||||||||||||     |||||||||  |||  ||||||||||||||||||||||
Db   43612 TCTGGGCAGTAGCTTTGGTCTACAGTCCATGNTCCCCGCCACGTCGCCGGTGGAGTCTGC 43671

Qy     710 gaac 713
           ||||
Db   43672 GAAC 43675


RESULT   7
AC009382
LOCUS       AC009382    92677 bp    DNA             HTG       18-FEB-2000
DEFINITION  Drosophila melanogaster clone RPCI98-1B21, *** SEQUENCING IN
            PROGRESS ***, 52 unordered pieces.
ACCESSION   AC009382
VERSION     AC009382.4  GI:6996682
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 92677)
  AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
            Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
            Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
            Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
            Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
            Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
            Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
            Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
            Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,
            Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 92677)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Feb 18, 2000 this sequence version replaced gi:5881494.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: DRAU
                Center clone name: RPCI98-1B21
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 42% of reads
                Chemistry: Dye-terminator Big Dye: 24% of reads
                Assembly program: Phrap; version 0.980611
```

```
    Consensus quality: 38748 bases at least Q40
    Consensus quality: 54037 bases at least Q30
    Consensus quality: 64416 bases at least Q20
    Estimated insert size: 135000; agarose-fp estimation
    Estimated insert size: 83602; sum-of-contigs estimation
    Quality coverage: 0.5x in Q20 bases; agarose-fp estimation
    Quality coverage: 0.7x in Q20 bases; sum-of-contigs estimation
--------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 52 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1      828: contig of 828 bp in length
*      829      848: gap of unknown length
*      849     2184: contig of 1336 bp in length
*     2185     2204: gap of unknown length
*     2205     3611: contig of 1407 bp in length
*     3612     3631: gap of unknown length
*     3632     4316: contig of 685 bp in length
*     4317     4336: gap of unknown length
*     4337     5322: contig of 986 bp in length
*     5323     5342: gap of unknown length
*     5343     6799: contig of 1457 bp in length
*     6800     6819: gap of unknown length
*     6820     7605: contig of 786 bp in length
*     7606     7625: gap of unknown length
*     7626     8704: contig of 1079 bp in length
*     8705     8724: gap of unknown length
*     8725     9498: contig of 774 bp in length
*     9499     9518: gap of unknown length
*     9519    10995: contig of 1477 bp in length
*    10996    11015: gap of unknown length
*    11016    11827: contig of 812 bp in length
*    11828    11847: gap of unknown length
*    11848    12755: contig of 908 bp in length
*    12756    12775: gap of unknown length
*    12776    13506: contig of 731 bp in length
*    13507    13526: gap of unknown length
*    13527    14778: contig of 1252 bp in length
*    14779    14798: gap of unknown length
*    14799    15619: contig of 821 bp in length
*    15620    15639: gap of unknown length
*    15640    17154: contig of 1515 bp in length
*    17155    17174: gap of unknown length
*    17175    18583: contig of 1409 bp in length
*    18584    18603: gap of unknown length
*    18604    19693: contig of 1090 bp in length
*    19694    19713: gap of unknown length
*    19714    21010: contig of 1297 bp in length
*    21011    21030: gap of unknown length
*    21031    22445: contig of 1415 bp in length
*    22446    22465: gap of unknown length
*    22466    23147: contig of 682 bp in length
*    23148    23167: gap of unknown length
*    23168    24468: contig of 1301 bp in length
*    24469    24488: gap of unknown length
*    24489    25609: contig of 1121 bp in length
*    25610    25629: gap of unknown length
*    25630    27332: contig of 1703 bp in length
*    27333    27352: gap of unknown length
*    27353    28358: contig of 1006 bp in length
*    28359    28378: gap of unknown length
*    28379    29948: contig of 1570 bp in length
*    29949    29968: gap of unknown length
*    29969    31673: contig of 1705 bp in length
*    31674    31693: gap of unknown length
*    31694    33056: contig of 1363 bp in length
*    33057    33076: gap of unknown length
*    33077    34658: contig of 1582 bp in length
*    34659    34678: gap of unknown length
*    34679    36591: contig of 1913 bp in length
*    36592    36611: gap of unknown length
*    36612    37617: contig of 1006 bp in length
*    37618    37637: gap of unknown length
*    37638    38903: contig of 1266 bp in length
*    38904    38923: gap of unknown length
*    38924    41024: contig of 2101 bp in length
*    41025    41044: gap of unknown length
*    41045    42943: contig of 1899 bp in length
*    42944    42963: gap of unknown length
*    42964    44680: contig of 1717 bp in length
*    44681    44700: gap of unknown length
*    44701    46683: contig of 1983 bp in length
*    46684    46703: gap of unknown length
*    46704    49058: contig of 2355 bp in length
*    49059    49078: gap of unknown length
*    49079    50931: contig of 1853 bp in length
*    50932    50951: gap of unknown length
*    50952    52258: contig of 1307 bp in length
*    52259    52278: gap of unknown length
*    52279    54271: contig of 1993 bp in length
*    54272    54291: gap of unknown length
*    54292    56574: contig of 2283 bp in length
*    56575    56594: gap of unknown length
*    56595    58695: contig of 2101 bp in length
*    58696    58715: gap of unknown length
*    58716    61026: contig of 2311 bp in length
*    61027    61046: gap of unknown length
*    61047    63594: contig of 2548 bp in length
*    63595    63614: gap of unknown length
*    63615    65977: contig of 2363 bp in length
*    65978    65997: gap of unknown length
*    65998    68802: contig of 2805 bp in length
*    68803    68822: gap of unknown length
*    68823    72229: contig of 3407 bp in length
*    72230    72249: gap of unknown length
*    72250    75163: contig of 2914 bp in length
*    75164    75183: gap of unknown length
*    75184    78054: contig of 2871 bp in length
*    78055    78074: gap of unknown length
*    78075    83202: contig of 5128 bp in length
*    83203    83222: gap of unknown length
*    83223    86848: contig of 3626 bp in length
*    86849    86868: gap of unknown length
*    86869    92677: contig of 5809 bp in length.
FEATURES             Location/Qualifiers
     source          1. .92677
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="RPCI98-1B21"
BASE COUNT    26406 a  19513 c  19116 g  26545 t   1097 others
ORIGIN


  Query Match            12.3%;  Score 462.8;  DB 51;  Length 92677;
  Best Local Similarity   98.7%;  Pred. No. 1.7e-99;
  Matches  519;  Conservative    0;  Mismatches    2;  Indels    5;  Gaps    5;

Qy    3244 ggagtctgcgatcacgcaacgcccactcgtccatcttctgcggcagcaaggatcggcggc 3303
           ||||||||||||||||||| |||||||||||||||||||| ||||||||||||||
Db   41059 GGAGTCTGCGATCACGCAACG-CCACTCGTCCATCTTCTGCGGCA-CAAGGATCGGCGGC 41116

Qy    3304 agcagaacttcgtgggcggagagcgggaacgggcttcggccatgtacgtggcctgcaagt 3363
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41117 A-CAGAACTTCGTGGGCGGAGAGCGGGAACGGGCTTCGGCCATGTACGTGGCCTGCAAGT 41175

Qy    3364 atctcccgcctgcgctgctcagctcccc-gggtgaacgtgctggcatgttagccgaggcg 3422
           |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   41176 ATCTCCCGCCTGCGCTGCTCAGCTCCCCNGGGTGAACGTGCTGGCATGTTAGCCGAGGCG 41235

Qy    3423 gccaagaccctggagaaggtgggcgacaagcgaaagctcaaggagtgctaccagctgatg 3482
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
```

```
Db  41236 GCCAAGACCCTGGAGAAGGTGGGCGACAAGCGAAA-CTCAAGGAGTGCTACCAGCTGATG 41294

Qy   3483 aagtcgctgggcaacggcattggcagcgtgaaggcttaggatagtagtgaagtacataat 3542
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db  41295 AAGTCGCTGGGCAACGGCATTGGCAGCGTGAAGGCTTACGATAGTAGTGAAGTACATAAT 41354

Qy   3543 aagtggcacgaacgtggtgtggattttcagcaaatgaatacccgtttgctattcaaaaga 3602
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  41355 AAGTGGCACGAACGTGGTGTGGATTTTCAGCAAATGAATACCCGTTTGCTATTCAAAAGA 41414

Qy   3603 attacaaatgcctaggtctttataattacgctattcctctgttttccacgcccggttatg 3662
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  41415 ATTACAAATGCCTAGGTCTTTATAATTACGCTATTCCTCTGTTTTCCACGCCCGGTTATG 41474

Qy   3663 cttagattgtaattttaaaattatttaatatggacattttatttgtttattatttaccgt 3722
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  41475 CTTAGATTGTAATTTTAAAATTATTTAATATGGACATTTTATTTGTTTATTATTTACCGT 41534

Qy   3723 acttgttaaacgtatttataacaataaatattttaacagatttaaa 3768
          |||||||||||||||||||||||||||||||||||||||||||| |
Db  41535 ACTTGTTAAACGTATTTATAACAATAAATATTTTAACAGATTTACA 41580


RESULT   8
U00968
LOCUS       U00968       4154 bp    mRNA            PRI       02-MAR-2000
DEFINITION  Human SREBP-1 mRNA, complete cds.
ACCESSION   U00968
VERSION     U00968.1  GI:409404
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Teleostomi;
            Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 4154)
  AUTHORS   Yokoyama,C., Wang,X., Briggs,M.R., Admon,A., Wu,J., Hua,X.,
            Goldstein,J.L. and Brown,M.S.
  TITLE     SREBP-1, a basic-helix-loop-helix-leucine zipper protein that
            controls transcription of the low density lipoprotein receptor gene
  JOURNAL   Cell 75 (1), 187-197 (1993)
  MEDLINE   94006541
REFERENCE   2  (bases 1 to 4154)
  AUTHORS   Goldstein,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-1993) Joseph Goldstein, Molecular Genetics,
            University of Texas Southwestern Medical Center at Dallas, 5323
            Harry Hines Blvd., Dallas, Texas, 75235, USA
FEATURES             Location/Qualifiers
     source          1. .4154
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="HeLa S3"
     CDS             167. .3610
                     /codon_start=1
                     /product="SREBP-1"
                     /protein_id="AAC50051.2"
                     /db_xref="GI:7144550"
                     /translation="MDEPPFSEAALEQALGEPCDLDAALLTDIEDMLQLINNQDSDFP
                     GLFDPPYAGSGAGGTDPASPDTSSPGSLSPPPATLSSSLEAFLSGPQAAPSPLSPPQP
                     APTPLKMYPSMPAFSPGPGIKEESVPLSILQTPTPQPLPGALLPQSFPAPAPPQFSST
                     PVLGYPSPPGGFSTGSPPGNTQQPLPGLPLASPPGVPPVSLHTQVQSVVPQQLLTVTA
                     APTAAPVTTTVTSQIQQVPVLLQPHFIKADSLLLTAMKTDGATVKAAGLSPLVSGTTV
                     QTGPLPTLVSGGTILATVPLVVDAEKLPINRLAAGSKAPASAQSRGEKRTAHNAIEKR
                     YRSSINDKIIELKDLVVGTEAKLNKSAVLRKAIDYIRFLQHSNQKLKQENLSLRTAVH
                     KSKSLKDLVSACGSGGNTDVLMEGVKTEVEDTLTPPPSDAGSPFQSSPLSLGSRGSGS
                     GGSGSDSEPDSPVFEDSKAKPEQRPSLHSRGMLDRSRLALCTLVFLCLSCNPLASLLG
                     ARGLPSPSDTTSVYHSPGRNVLGTESRDGPGWAQWLLPPVVWLLNGLLVLVSLVLLFV
                     YGEPVTRPHSGPAVYFWRHRKQADLDLARGDFAQAAQQLWLALRALGRPLPTSHLDLA
                     CSLLWNLIRHLLQRLWVGRWLAGRAGGLQQDCALRVDASASARDAALVYHKLHQLHTM
                     GKHTGGHLTATNLALSALNLAECAGDAVSVATLAEIYVAAALRVKTSLPRALHFLTRF
                     FLSSARQACLAQSGSVPPAMQWLCHPVGHRFFVDGDWSVLSTPWESLYSLAGNPVDPL
                     AQVTQLFREHLLERALNCVTQPNPSPGSADGDKEFSDALGYLQLLNSCSDAAGAPAYS
                     FSISSSMATTTGVDPVAKWWASLTAVVIHWLRRDEEAAERLCPLVEHLPRVLQESERP
                     LPRAALHSFKAARALLGCAKAESGPASLTICEKASGYLQDSLATTPASSSIDKAVQLF
                     LCDLLLVVRTSLWRQQQPPAPAPAAQGASSRPQASALELRGFQRDLSSLRRLAQSFRP
                     AMRRVFLHEATARLMAGASPTRTHQLLDRSLRRRAGPGGKGGAVAELEPRPTRREHAE
                     ALLLASCYLPPGFLSAPGQRVGMLAEAARTLEKLGDRRLLHDCQQMLMRLGGGTTVTS
                     S"
BASE COUNT      713 a   1406 c   1281 g    754 t
ORIGIN


  Query Match             4.4%;  Score 166.6;  DB 50;  Length 4154;
  Best Local Similarity   45.8%;  Pred. No. 6e-29;
  Matches 1145;  Conservative    0;  Mismatches 1294;  Indels   62;  Gaps   14;

Qy   1023 gaagtaaagcgctcggcccacaacgccatcgagcggcgctatcgcacctcaatcaacgac 1082
          | ||  |||||| | |||||||||||||| |||  |||||| ||| |||| ||||| |||
Db   1130 GGAGAGAAGCGCACAGCCCACAACGCCATTGAGAAGCGCTACCGCTCCTCCATCAATGAC 1189

Qy   1083 aagattaacgagttgaagaacttggtagtgggagagcaggccaagctgaacaagtccgca 1142
          || || |  ||| | ||| |  |||| |||||     |||| |||||||| || || ||
Db   1190 AAAATCATTGAGCTCAAGGATCTGGTGGTGGGCACTGAGGCAAAGCTGAATAAATCTGCT 1249

Qy   1143 gtgttgcggaaatccatagacaagattcgggatctgcaacgccagaatcacgatctgaag 1202
          || ||||| ||  |||| ||| | |||||   |||||||||   || ||  | || |||
Db   1250 GTCTTGCGCAAGGCCATCGACTACATTCGCTTTCTGCAACACAGCAACCAGAAACTCAAG 1309

Qy   1203 gcagagttgcagcgcctgcagagggagctaatggcacgcgacggctccaaggtgaaggat 1262
            |||   |  | ||||  |  |  |      | || |     ||||    ||
Db   1310 CAGGAGAACCTAAGTCTGCGCACTGCTGTCCACAAAAGCAAATCTCTGAAGGATCTGGTG 1369

Qy   1263 t----tacttcagctgggcactcggcctggtagagcatccaagaagcgccgcgagagctc 1318
          |    |     |||        |     |||  | |  | |  |    ||| |
Db   1370 TCGGCCTGTGGCAGTGGAGGGAACACAGACGTGCTCATGGAGGGCGTGAAGACTGAGGTG 1429

Qy   1319 gcagacctttaccacggatgccggactgacgccgccacgcagcgatgaatcggatccttc 1378
          |  |||       | |  ||   | || || | | |       | | |  ||||
Db   1430 GAGGACACACTGACCCCACCCCCCTCGGATGCTGGCTCACCTTTCCAGAGCAGCCCCTTG 1489

Qy   1379 gctctcgcccatgcactcggacatctcgttgccgccatcacctatggtggatccaccgc 1438
          | || | |       | |      || || |     ||  |  | |  |    || ||
Db   1490 TCCCTTGGCAGCAGGGGCAGTGGCAGCGGTGGCAGTGGCAGTGACTCGGAGCCTGACAGC 1549

Qy   1439 cagctgtagcagtggcagcagcagcagcaatgaagaaccactggtggtgcccagctctat 1498
          |   | |   || | ||||     |     ||  | |  | |    ||| ||||
Db   1550 CCAGTCTTTGAG-GACAGCAAGGCAAAGCCAGAGCAGCGGCCGTCTCTGCACAGC----- 1603

Qy   1499 gcgcggcatggccacccactctcgcctcggactctgcatgtttatgttcgccatcctggc 1558
           || ||||||     || ||| ||||| |  || |||| | ||  | |||  |  |||| |
Db   1604 -CGGGGCATGCTGGACCGCTCCCGCCTGGCCCTGTGCACGCTCGTCTTCCTCTGCCTGTC 1662

Qy   1559 cgtcaatcccttcaagacctttctccagcgcggccactatgacagtaatgacgatcttgg 1618
          |  ||| |||||     |||| ||   | |  |       | | |||       |||
Db   1663 CTGCAACCCCTTGGCCTCCTTGCTGGGGGCCCGGGGGCTTCCCAGCCCCTCAGATACCAC 1722

Qy   1619 cgacatgagc---------ggtcaaagacgcattctctcttacgacgtggaaggtgaagg 1669
          |  | |   |         || |  |  |  | | ||   |   |  || ||   ||
Db   1723 CAGCGTCTACCATAGCCCTGGGCGCAACGTGCTGGGCACCGAGAGCAGAGATGGCCCTGG 1782

Qy   1670 ttttgctgtctggcagcagagttcctggatatggctattgaacttcacactgatgcttgg 1729
           |  ||    |||| || |    |   | | |||||  | ||         || |||| |
Db   1783 CTGGGCCCAGTGGCTGCTGCCCCCAGTGGTCTGGCTGCTCAATGGGCTGTTGGTGCTCGT 1842

Qy   1730 atgcttggtgaaattgctggtttacggtgatcc--------gcagctggacgcgcaaac 1780
          | |||||||    | | || ||||||||  ||        ||  |     |  |
Db   1843 CTCCTTGGTGCTTCTCTTTGTCTACGGTGAGCCAGTCACACGGCCCCACTCAGGCCCCGC 1902

Qy   1781 ggacgcctactgccagcacaggcagcgggctgacttctatttagccaaggacagtcgtc 1840
          |    || |||   |||  |  ||| |||||||  |  |  |    ||  ||| | |  |
Db   1903 CGTGTACTTCTGGAGGCATCGCAAGCAGGCTGACCTGGACCTGGCCCGGGGAGACTTTGC 1962

Qy   1841 tcaggcctacgccggttacctcaactgtctgcatatgtttggattaagtctaccggcgtc 1900
          |||||   |                  ||||      | ||       || ||  | ||
Db   1963 CCAGGCTGCCCAGCAGCTGTGGCTGGCCCTGCGGGCACTGGGCCGGCCCCTGCCCACCTC 2022
```

```
Qy   1901 gcgcttggagtgttacttgcagaccacgtggcagttccttcgttttcttttccatcgcct 1960
Db   2023 CCACCTGGACCTGGCTTGTAGCCTCCTCTGGAACCTCATCCGTCACCTGCTGCAGCGTCT 2082

Qy   1961 ctggctgggtcgggtgctgtcacggcggtccggtgggctgtttagcaacgccgccagcag 2020
Db   2083 CTGGGTGGGCCGCTGGCTGGCAGGCCGGGCAGGGGGCCTGCAGCAGGACTGTGCTCTGCG 2142

Qy   2021 gaaacaggcgctggcatctgcacgcgaactggccctgctcttcaaccgactgaatcaatt 2080
Db   2143 AGTGGATGCTAGCGCCAGCGCCCGAGACGCAGCCCTGGTCTACCATAAGCTGCACCAGCT 2202

Qy   2081 gcaactgactggaaatggaagccgcggtgacatgaacggcattatgatggcactattcgc 2140
Db   2203 GCACACCATGGGGAAGCACACAGGCGGGCACCTCACTGCCACCAACCTGGCGCTGAGTGC 2262

Qy   2141 aagcaacatggctgaagtggcgcacaatctactgacaccgcgcgagaccatctgcatcca 2200
Db   2263 CCTGAACCTGGCAGAGTGTGCAGGGGATGCCGTGTCTGTGGCGACGCTGGCCGAGATCTA 2322

Qy   2201 cgtaatgacagcgttgcgaatgaagcgcagtgccccaaaatggttgcaacagttcttcgc 2260
Db   2323 TGTGGCGGCTGCATTGAGAGTGAAGACCAGTCTCCCACGGGCCTTG---CATTTTCTGAC 2379

Qy   2261 ccgatactacatgagccgggctcgtcaagagtgcggtcgcactagggccaccgagcaaac 2320
Db   2380 ACGCTTCTTCCTGAGCAGTGCCCGCCAGGCCTGCCTGGCACAGAGTGGCTCAGTGC---C 2436

Qy   2321 gcaggagctacgttgggcattcacagcctatggatatcgctactgcgccacgcacgtctt 2380
Db   2437 TCCTGCCATGCAGTGGCTCTGCCACCCCGTGGGCCACCGTTTCTTCG---TGGATGGGGA 2493

Qy   2381 cacgtacgatctgagcgactccggcgagcaggatggattcttcacacgtcttaggaatcc 2440
Db   2494 CTGGTCCGTGCTCAGTACCCCATG------GGAGAGCCTGTACAGCTTGGCCGGGAACCC 2547

Qy   2441 atgtgatcccgctgcccacgtcattaagcaatatcgagagcatttgctgtttaaatccat 2500
Db   2548 AGTGGACCCCCTGGCCCAGGTGACTCAGCTATTCCGGGAACATCTCTTAGAGCGAGCACT 2607

Qy   2501 tcagtgtctggtaggagcgggccacaaatcgggaggcctgcccacatcttctgtcagcgg 2560
Db   2608 GAACTGTGTGACCCAGCCCAACCCCAGCCCTGGGTCAGCTGATGGGGACAAGGAATTCTC 2667

Qy   2561 agaggcggaacagttgcagcaacagcagcacagcggcaccattgtcagcaatgttcttaa 2620
Db   2668 GGATGCCCTCGGGTACCTGCAGCTGCTGAACAGCTGTTCTGATGCTGCGGGGGCTCCTGC 2727

Qy   2621 gtacacgtccctccttaaggacactctctgggctgatgaggatgagcgggatacaaacgt 2680
Db   2728 CTACAGCTTCTCCATCAGTTCCA---GCATGGCCACCACCACCGGCGTAGACCCGGTGGC 2784

Qy   2681 ggtgtggtgggccgatgttttggagaccgcagtgcactggctccttggtgaagacacgct 2740
Db   2785 CAAGTGGTGGGCCTCTCTGACAGCTGTGGTGATCCACTGGCTGCGGCGGGATGAGGAGGC 2844

Qy   2741 ggccgagcaattgtacggcaggatcaagcaaatgcccacgcagctgcaacagtgcggcga 2800
Db   2845 GGCTGAGCGGCTGTGCCCGCTGGTGGAGCACCTGCCCCGGGTGCTGCAGGAGTCTGAGAG 2904

Qy   2801 aaacgatcatctgcccaaggcgctgcatgctgtgctgcgagctaagatgatcttactaaa 2860
Db   2905 ACCCCTGCCCAGGGC---AGCTCTGCACTCCTTCAAGGCTGCCCGGGCCCTGCTGGGCTG 2961

Qy   2861 aaacaatggcaacgcactggacaaaagtctcaagcaattggtaaacatcctctgcgatga 2920
Db   2962 TGCCAA-GGCAGAGTCTGGTCCAGCCAGCCTGACCATCTGTGAGAAGGCCAGTGGGTACC 3020

Qy   2921 gtcgagtgtggagctccaagagtgcttgactgtcaaccggatcaccgacgccaagggtat 2980
Db   3021 TGCAGGACAGCCTGGCTACCACACCAG--CCAGCAGCTCCATTGACAAGGCCGTGCAGCT 3078

Qy   2981 aaagctgcttttccagttgcttacctgcgattggctgctcgaaactaggactgctctgtg 3040
Db   3079 GTTCCTGTGTGACCTGCTTCTTGTGGTGCGCACCAGCCTGTGGCGGCAGCAGCAGCCCCC 3138

Qy   3041 ggaactggaacacatgaatatggaggacgatggcttctaccaagtgccaggtgaagtgct 3100
Db   3139 GGCCCCGGCCCCAGCAGCCCAGGGCGCCAGCAGCAGGCCCCAGGCTTCCGCCCTTGAGCT 3198

Qy   3101 cgagaagttccagaccgatttgaactcgttgcgcaacattgtggagaatataccgaacgc 3160
Db   3199 GCGTGGCTTCCAACGGGACCTGAGCAGCCTGAGGCGGCTGGCACAGAGCTTCCGGCCCGC 3258

Qy   3161 ccaatcgcgcatatatttgtacgaggcagtttgtcgcctgatggctggagcctcaccgtg 3220
Db   3259 CATGCGGAGGGTGTTCCTACATGAGGCCACGGCCCGGCTGATGGCGGGGGCCAGCCCCAC 3318

Qy   3221 tccaacgcaacagctcttggacaggagtctgcgatcacgcaacgcccactcgtccatctt 3280
Db   3319 ACGGACACACCAGCTCCTCGACCGCAGTCTGAGGCGGCGGGCAGGCCCCGGTGGCAAAGG 3378

Qy   3281 ctgcggcagcaaggatcggcggcagcagaacttcgtgggcggagagcgggaacgggcttc 3340
Db   3379 AGGCGCGGTGGCGGAGCTGGAGCCGC---------GGCCCACGCGGCGGGAGCACGCGGA 3429

Qy   3341 ggccatgtacgtggcctgcaagtatctcccgcctgcgctgctcagctccccgggtgaacg 3400
Db   3430 GGCCTTGCTGCTGGCCTCCTGCTACCTGCCCCCCGGCTTCCTGTCGGCGCCCGGGCAGCG 3489

Qy   3401 tgctggcatgttagccgaggcggccaagaccctggagaaggtgggcgacaagcgaaagct 3460
Db   3490 CGTGGGCATGCTGGCTGAGGCGGCGCGCACACTCGAGAAGCTTGGCGATCGCCGGCTGCT 3549

Qy   3461 caaggagtgctaccagctgatgaaagtcgctgggcaacggca 3501
Db   3550 GCACGACTGTCAGCAGATGCTCATGCGCCTGGGCGGTGGGA 3590

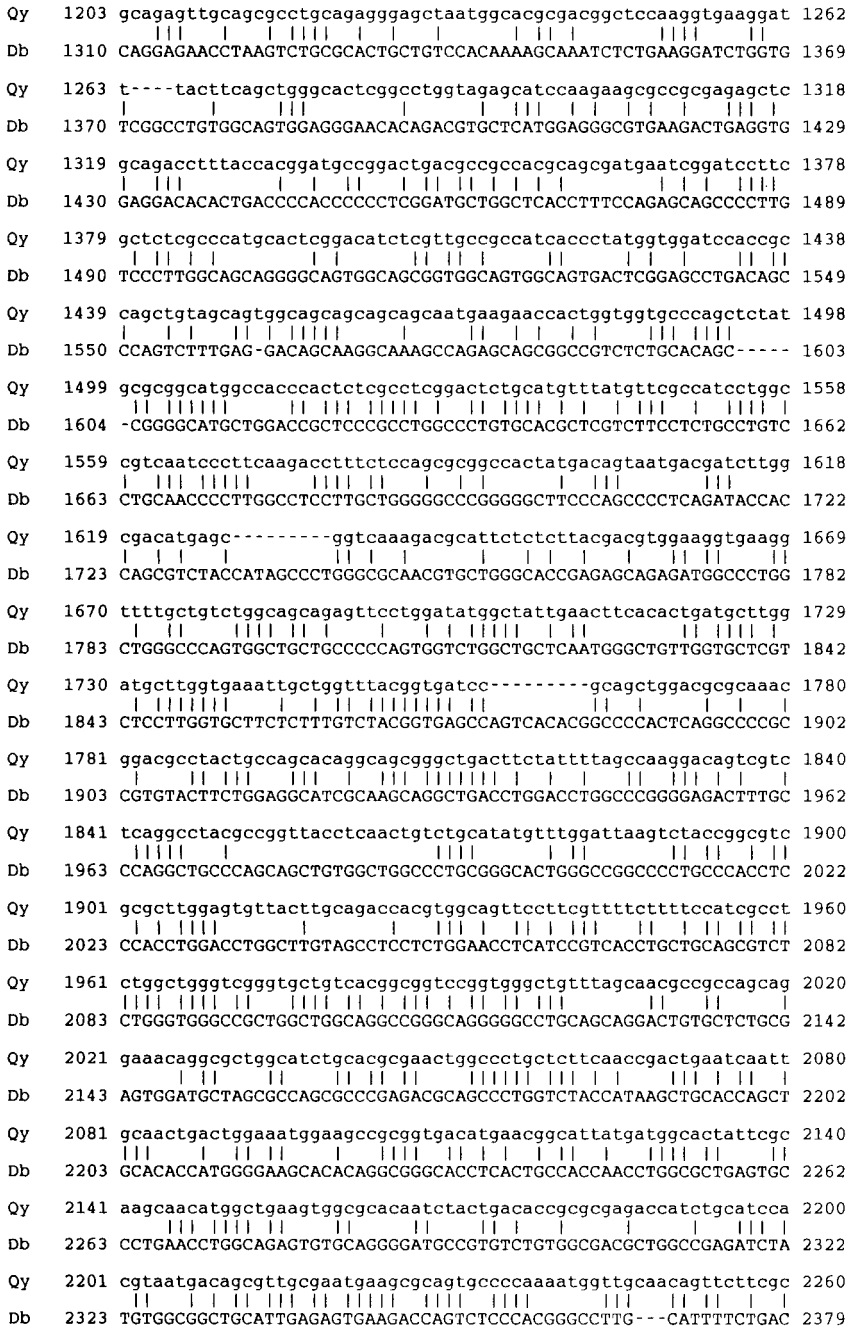

RESULT   9
AR069298
LOCUS       AR069298     4154 bp    DNA             PAT       18-FEB-2000
DEFINITION  Sequence 37 from patent US 5891631.
ACCESSION   AR069298
VERSION     AR069298.1  GI:7220186
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4154)
  AUTHORS   Goldstein,J.L., Brown,M.S., Briggs,M.R. and Wang,X.
  TITLE     Methods relating tosterol regulatory element binding proteins
  JOURNAL   Patent: US 5891631-A 37 06-APR-1999;
FEATURES             Location/Qualifiers
     source          1. .4154
                     /organism="unknown"
BASE COUNT      714 a   1406 c   1280 g    754 t
ORIGIN


  Query Match             4.4%;  Score 165;  DB 5;  Length 4154;
  Best Local Similarity   45.7%;  Pred. No. 1.4e-28;
  Matches 1144;  Conservative    0;  Mismatches 1295;  Indels   62;  Gaps   14;

Qy   1023 gaagtaaagcgctcggcccacaacgccatcgagcggcgctatcgcacctcaatcaacgac 1082
Db   1130 GGAGAGAAGCGCACAGCCCACAACGCCATTGAGAAGCGCTACCGCTCCTCCATCAATGAC 1189

Qy   1083 aagattaacgagttgaagaacttggtagtgggagagcaggccaagctgaacaagtccgca 1142
Db   1190 AAAATCATTGAGCTCAAGGATCTGGTGGTGGGCACTGAGGCAAAGCTGAATAAATCTGCT 1249

Qy   1143 gtgttgcggaaatccatagacaagattcgggatctgcaacgccagaatcacgatctgaag 1202
Db   1250 GTCTTGCGCAAGGCCATCGACTACATTCGCTTTCTGCAACACAGCAACCAGAAACTCAAG 1309
```


Qy 1203 gcagagttgcagcgcctgcagagggagctaatggcacgcgacggctccaaggtgaaggat 1262
Db 1310 CAGGAGAACCTAAGTCTGCGCACTGCTGTCCACAAAAGCAAATCTCTGAAGGATCTGGTG 1369
Qy 1263 t----tacttcagctgggcactcggcctggtagagcatccaagaagcgccgcgagagctc 1318
Db 1370 TCGGCCTGTGGCAGTGGAGGGAACACAGACGTGCTCATGGAGGGCGTGAAGACTGAGGTG 1429
Qy 1319 gcagacctttaccacggatgccggactgacgccgccacgcagcgatgaatcggatccttc 1378
Db 1430 GAGGACACACTGACCCCACCCCCCTCGGATGCTGGCTCACCTTTCCAGAGCAGCCCCTTG 1489
Qy 1379 gctctcgcccatgcactcggacatctcgttgccgccatcaccctatggtggatccaccgc 1438
Db 1490 TCCCTTGGCAGCAGGGGCAGTGGCAGCGGTGGCAGTGGCAGTGACTCGGAGCCTGACAGC 1549
Qy 1439 cagctgtagcagtggcagcagcagcagcaatgaagaaccactggtggtgcccagctctat 1498
Db 1550 CCAGTCTTTGAG-GACAGCAAGGCAAAGCCAGAGCAGCGGCCGTCTCTGCACAGC----- 1603
Qy 1499 gcgcggcatggccacccactctcgcctcggactctgcatgtttatgttcgccatcctggc 1558
Db 1604 -CGGGGCATGCTGGACCGCTCCCGCCTGGCCCTGTGCACGCTCGTCTTCCTCTGCCTGTC 1662
Qy 1559 cgtcaatcccttcaagacctttctccagcgcggccactatgacagtaatgacgatcttgg 1618
Db 1663 CTGCAACCCCTTGGCCTCCTTGCTGGGGGCCCGGGGGCTTCCCAGCCCCTCAGATACCAC 1722
Qy 1619 cgacatgagc---------ggtcaaagacgcattctctcttacgacgtggaaggtgaagg 1669
Db 1723 CAGCGTCTACCATAGCCCTGGGCGCAACGTGCTGGGCACCGAGAGCAGAGATGGCCCTGG 1782
Qy 1670 ttttgctgtctggcagcagagttcctggatatggctattgaacttcacactgatgcttgg 1729
Db 1783 CTGGGCCCAGTGGCTGCTGCCCCCAGTGGTCTGGCTGCTCAATGGGCTGTTGGTGCTCGT 1842
Qy 1730 atgcttggtgaaattgctggtttacggtgatcc---------gcagctggacgcgcaaac 1780
Db 1843 CTCCTTGGTGCTTCTCTTTGTCTACGGTGAGCCAGTCACACGGCCCCACTCAGGCCCCGC 1902
Qy 1781 ggacgcctactgccagcacaggcagcgggctgacttctatttttagccaaggacagtcgtc 1840
Db 1903 CGTGTACTTCTGGAGGCATCGCAAGCAGGCTGACCTGGACCTGGCCCGGGGAGACTTTGC 1962
Qy 1841 tcaggcctacgccggttacctcaactgtctgcatatgtttggattaagtctaccggcgtc 1900
Db 1963 CCAGGCTGCCCAGCAGCTGTGGCTGGCCCTGCGGGCACTGGGCCGGCCCCTGCCCACCTC 2022
Qy 1901 gcgcttggagtgttacttgcagaccacgtggcagttccttcgttttcttttccatcgcct 1960
Db 2023 CCACCTGGACCTGGCTTGTAGCCTCCTCTGGAACCTCATCCGTCACCTGCTGCAGCGTCT 2082
Qy 1961 ctggctgggtcgggtgctgtcacggcggtccggtgggctgtttagcaacgccgccagcag 2020
Db 2083 CTGGGTGGGCCGCTGGCTGGCAGGCCGGGCAGGGGGCCTGCAGCAGGACTGTGCTCTGCG 2142
Qy 2021 gaaacaggcgctggcatctgcacgcgaactggccctgctcttcaaccgactgaatcaatt 2080
Db 2143 AGTGGATGCTAGCGCCAGCGCCCGAGACGCAGCCCTGGTCTACCATAAGCTGCACCAGCT 2202
Qy 2081 gcaactgactggaaatggaagccgcggtgacatgaacggcattatgatggcactattcgc 2140
Db 2203 GCACACCATGGGGAAGCACACAGGCGGGCACCTCACTGCCACCAACCTGGCGCTGAGTGC 2262
Qy 2141 aagcaacatggctgaagtggcgcacaatctactgacaccgcgcgagaccatctgcatcca 2200
Db 2263 CCTGAACCTGGCAGAGTGTGCAGGGGATGCCGTGTCTGTGGCGACGCTGGCCGAGATCTA 2322
Qy 2201 cgtaatgacagcgttgcgaatgaagcgcagtgccccaaaatggttgcaacagttcttcgc 2260
Db 2323 TGTGGCGGCTGCATTGAGAGTGAAGACCAGTCTCCCACGGGCCTTG---CATTTTCTGAC 2379
Qy 2261 ccgatactacatgagccgggctcgtcaagagtgcggtcgcactagggccaccgagcaaac 2320
Db 2380 ACGCTTCTTCCTGAGCAGTGCCCGCCAGGCCTGCCTGGCACAGAGTGGCTCAGTGC---C 2436
Qy 2321 gcaggagctacgttgggcattcacagcctatggatatcgctactgcgccacgcacgtctt 2380
Db 2437 TCCTGCCATGCAGTGGCTCTGCCACCCCGTGGGCCACCGTTTCTTCG---TGGATGGGGA 2493
Qy 2381 cacgtacgatctgagcgactccggcgagcaggatggattcttcacacgtcttaggaatcc 2440
Db 2494 CTGGTCCGTGCTCAGTACCCCATG------GGAGAGCCTGTACAGCTTGGCCGGGAACCC 2547
Qy 2441 atgtgatcccgctgcccacgtcattaagcaatatcgagagcatttgctgtttaaatccat 2500
Db 2548 AGTGGACCCCCTGGCCCAGGTGACTCAGCTATTCCGGGAACATCTCTTAGAGCGAGCACT 2607
Qy 2501 tcagtgtctggtaggagcgggccacaaatcgggaggcctgcccacatcttctgtcagcgg 2560
Db 2608 GAACTGTGTGACCCAGCCCAACCCCAGCCCTGGGTCAGCTGATGGGGACAAGGAATTCTC 2667
Qy 2561 agaggcggaacagttgcagcaacagcagcacagcggcaccattgtcagcaatgttcttaa 2620
Db 2668 GGATGCCCTCGGGTACCTGCAGCTGCTGAACAGCTGTTCTGATGCTGCGGGGGCTCCTGC 2727
Qy 2621 gtacacgtccctccttaaggacactctctgggctgatgaggatgagcgggatacaaacgt 2680
Db 2728 CTACAGCTTCTCCATCAGTTCCA---GCATGGCCACCACCACCGGCGTAGACCCGGTGGC 2784
Qy 2681 ggtgtggtgggccgatgttttggagaccgcagtgcactggctccttggtgaagacacgct 2740
Db 2785 CAAGTGGTGGGCCTCTCTGACAGCTGTGGTGATCCACTGGCTGCGGCGGGATGAGGAGGC 2844
Qy 2741 ggccgagcaattgtacggcaggatcaagcaaatgcccacgcagctgcaacagtgcggcga 2800
Db 2845 GGCTGAGCGGCTGTGCCCGCTGGTGGAGCACCTGCCCCGGGTGCTGCAGGAGTCTGAGAG 2904
Qy 2801 aaacgatcatctgcccaaggcgctgcatgctgtgctgcgagctaagatgatcttactaaa 2860
Db 2905 ACCCCTGCCCAGGGC---AGCTCTGCACTCCTTCAAGGCTGCCCGGGCCCTGCTGGGCTG 2961
Qy 2861 aaacaatggcaacgcactggacaaaagtctcaagcaattggtaaacatcctctgcgatga 2920
Db 2962 TGCCAA-GGCAGAGTCTGGTCCAGCCAGCCTGACCATCTGTGAGAAGGCCAGTGGGTACC 3020
Qy 2921 gtcgagtgtggagctccaagagtgcttgactgtcaaccggatcaccgacgccaagggtat 2980
Db 3021 TGCAGGACAGCCTGGCTACCACACCAG--CCAGCAGCTCCATTGACAAGGCCGTGCAGCT 3078
Qy 2981 aaagctgcttttccagttgcttacctgcgattggctgctcgaaactaggactgctctgtg 3040
Db 3079 GTTCCTGTGTGACCTGCTTCTTGTGGTGCGCACCAGCCTGTGGCGGCAGCAGCAGCCCCC 3138
Qy 3041 ggaactggaacacatgaatatggaggacgatggcttctaccaagtgccaggtgaagtgct 3100
Db 3139 GGCCCCGGCCCCAGCAGCCCAGGGCACCAGCAGCAGGCCCCAGGCTTCCGCCCTTGAGCT 3198
Qy 3101 cgagaagttccagaccgatttgaactcgttgcgcaacattgtggagaatataccgaacgc 3160
Db 3199 GCGTGGCTTCCAACGGGACCTGAGCAGCCTGAGGCGGCTGGCACAGAGCTTCCGGCCCGC 3258
Qy 3161 ccaatcgcgcatatatttgtacgaggcagtttgtcgcctgatggctggagcctcaccgtg 3220
Db 3259 CATGCGGAGGGTGTTCCTACATGAGGCCACGGCCCGGCTGATGGCGGGGGCCAGCCCCAC 3318
Qy 3221 tccaacgcaacagctcttggacaggagtctgcgatcacgcaacgcccactcgtccatctt 3280
Db 3319 ACGGACACACCAGCTCCTCGACCGCAGTCTGAGGCGGCGGGCAGGCCCCGGTGGCAAAGG 3378
Qy 3281 ctgcggcagcaaggatcggcggcagcagaacttcgtgggcggagagcgggaacgggcttc 3340
Db 3379 AGGCGCGGTGGCGGAGCTGGAGCCGC---------GGCCCACGCGGCGGGAGCACGCGGA 3429
Qy 3341 ggccatgtacgtggcctgcaagtatctcccgcctgcgctgctcagctccccgggtgaacg 3400

```
Db   3430 GGCCTTGCTGCTGGCCTCCTGCTACCTGCCCCCCGGCTTCCTGTCGGCGCCCGGGCAGCG 3489

Qy   3401 tgctggcatgttagccgaggcggccaagaccctggagaaggtgggcgacaagcgaaagct 3460

Db   3490 CGTGGGCATGCTGGCTGAGGCGGCGCGCACACTCGAGAAGCTTGGCGATCGCCGGCTGCT 3549

Qy   3461 caaggagtgctaccagctgatgaagtcgctgggcaacggca 3501

Db   3550 GCACGACTGTCAGCAGATGCTCATGCGCCTGGGCGGTGGGA 3590


RESULT  10
I22310
LOCUS       I22310       4154 bp    DNA             PAT       07-OCT-1996
DEFINITION  Sequence 37 from patent US 5527690.
ACCESSION   I22310
VERSION     I22310.1  GI:1602664
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4154)
  AUTHORS   Goldstein,J.L., Brown,M.S., Briggs,M.R. and Wang,X.
  TITLE     Methods and compositions relating to sterol regulatory element
            binding proteins
  JOURNAL   Patent: US 5527690-A 37 18-JUN-1996;
FEATURES             Location/Qualifiers
     source          1. .4154
                     /organism="unknown"
BASE COUNT      714 a   1406 c   1280 g    754 t
ORIGIN


  Query Match             4.4%;  Score 165;  DB 5;  Length 4154;
  Best Local Similarity   45.7%;  Pred. No. 1.4e-28;
  Matches 1144;  Conservative    0;  Mismatches 1295;  Indels   62;  Gaps   14;

Qy   1023 gaagtaaagcgctcggcccacaacgccatcgagcggcgctatcgcacctcaatcaacgac 1082

Db   1130 GGAGAGAAGCGCACAGCCCACAACGCCATTGAGAAGCGCTACCGCTCCTCCATCAATGAC 1189

Qy   1083 aagattaacgagttgaagaacttggtagtgggagagcaggccaagctgaacaagtccgca 1142

Db   1190 AAAATCATTGAGCTCAAGGATCTGGTGGTGGGCACTGAGGCAAAGCTGAATAAATCTGCT 1249

Qy   1143 gtgttgcggaaatccatagacaagattcgggatctgcaacgccagaatcacgatctgaag 1202

Db   1250 GTCTTGCGCAAGGCCATCGACTACATTCGCTTTCTGCAACACAGCAACCAGAAACTCAAG 1309

Qy   1203 gcagagttgcagcgcctgcagagggagctaatggcacgcgacggctccaaggtgaaggat 1262

Db   1310 CAGGAGAACCTAAGTCTGCGCACTGCTGTCCACAAAAGCAAATCTCTGAAGGATCTGGTG 1369

Qy   1263 t----tacttcagctgggcactcggcctggtagagcatccaagaagcgccgcgagagctc 1318

Db   1370 TCGGCCTGTGGCAGTGGAGGGAACACAGACGTGCTCATGGAGGGCGTGAAGACTGAGGTG 1429

Qy   1319 gcagacctttaccacggatgccggactgacgccgccacgcagcgatgaatcggatccttc 1378

Db   1430 GAGGACACACTGACCCCACCCCCCTCGGATGCTGGCTCACCTTTCCAGAGCAGCCCCTTG 1489

Qy   1379 gctctcgcccatgcactcggacatctcgttgccgccatcaccctatggtggatccaccgc 1438

Db   1490 TCCCTTGGCAGCAGGGGCAGTGGCAGCGGTGGCAGTGGCAGTGACTCGGAGCCTGACAGC 1549

Qy   1439 cagctgtagcagtggcagcagcagcagcaatgaagaaccactggtggtgcccagctctat 1498

Db   1550 CCAGTCTTTGAG-GACAGCAAGGCAAAGCCAGAGCAGCGGCCGTCTCTGCACAGC----- 1603

Qy   1499 gcgcggcatggccacccactctcgcctcggactctgcatgtttatgttcgccatcctggc 1558

Db   1604 -CGGGGCATGCTGGACCGCTCCCGCCTGGCCCTGTGCACGCTCGTCTTCCTCTGCCTGTC 1662

Qy   1559 cgtcaatcccttcaagacctttctccagcgcggccactatgacagtaatgacgatcttgg 1618

Db   1663 CTGCAACCCCTTGGCCTCCTTGCTGGGGGCCCGGGGGCTTCCCAGCCCCTCAGATACCAC 1722

Qy   1619 cgacatgagc---------ggtcaaagacgcattctctcttacgacgtggaaggtgaagg 1669

Db   1723 CAGCGTCTACCATAGCCCTGGGCGCAACGTGCTGGGCACCGAGAGCAGAGATGGCCCTGG 1782

Qy   1670 ttttgctgtctggcagcagagttcctggatatggctattgaacttcacactgatgcttgg 1729

Db   1783 CTGGGCCCAGTGGCTGCTGCCCCCAGTGGTCTGGCTGCTCAATGGGCTGTTGGTGCTCGT 1842

Qy   1730 atgcttggtgaaattgctggtttacggtgatcc---------gcagctggacgcgcaaac 1780

Db   1843 CTCCTTGGTGCTTCTCTTTGTCTACGGTGAGCCAGTCACACGGCCCCACTCAGGCCCCGC 1902

Qy   1781 ggacgcctactgccagcacaggcagcgggctgacttctattttagccaaggacagtcgtc 1840

Db   1903 CGTGTACTTCTGGAGGCATCGCAAGCAGGCTGACCTGGACCTGGCCCGGGGAGACTTTGC 1962

Qy   1841 tcaggcctacgccggttacctcaactgtctgcatatgtttggattaagtctaccggcgtc 1900

Db   1963 CCAGGCTGCCCAGCAGCTGTGGCTGGCCCTGCGGGCACTGGGCCGGCCCCTGCCCACCTC 2022

Qy   1901 gcgcttggagtgttacttgcagaccacgtggcagttccttcgttttctttccatcgcct 1960

Db   2023 CCACCTGGACCTGGCTTGTAGCCTCCTCTGGAACCTCATCCGTCACCTGCTGCAGCGTCT 2082

Qy   1961 ctggctgggtcgggtgctgtcacggcggtccggtgggctgtttagcaacgccgccagcag 2020

Db   2083 CTGGGTGGGCCGCTGGCTGGCAGGCCGGGCAGGGGGCCTGCAGCAGGACTGTGCTCTGCG 2142

Qy   2021 gaaacaggcgctggcatctgcacgcgaactggccctgctcttcaaccgactgaatcaatt 2080

Db   2143 AGTGGATGCTAGCGCCAGCGCCCGAGACGCAGCCCTGGTCTACCATAAGCTGCACCAGCT 2202

Qy   2081 gcaactgactggaaatggaagccgcggtgacatgaacggcattatgatggcactattcgc 2140

Db   2203 GCACACCATGGGGAAGCACACAGGCGGGCACCTCACTGCCACCAACCTGGCGCTGAGTGC 2262

Qy   2141 aagcaacatggctgaagtggcgcacaatctactgacaccgcgcgagaccatctgcatcca 2200

Db   2263 CCTGAACCTGGCAGAGTGTGCAGGGGATGCCGTGTCTGTGGCGACGCTGGCCGAGATCTA 2322

Qy   2201 cgtaatgacagcgttgcgaatgaagcgcagtgccccaaaatggttgcaacagttcttcgc 2260

Db   2323 TGTGGCGGCTGCATTGAGAGTGAAGACCAGTCTCCCACGGGCCTTG---CATTTTCTGAC 2379

Qy   2261 ccgatactacatgagccgggctcgtcaagagtgcggtcgcactagggccaccgagcaaac 2320

Db   2380 ACGCTTCTTCCTGAGCAGTGCCCGCCAGGCCTGCCTGGCACAGAGTGGCTCAGTGC---C 2436

Qy   2321 gcaggagctacgttgggcattcacagcctatggatatcgctactgcgccacgcacgtctt 2380

Db   2437 TCCTGCCATGCAGTGGCTCTGCCACCCCGTGGGCCACCGTTTCTTCG---TGGATGGGGA 2493

Qy   2381 cacgtacgatctgagcgactccggcgagcaggatggattcttcacacgtcttaggaatcc 2440

Db   2494 CTGGTCCGTGCTCAGTACCCCATG------GGAGAGCCTGTACAGCTTGGCCGGGAACCC 2547

Qy   2441 atgtgatcccgctgcccacgtcattaagcaatatcgagagcatttgctgtttaaatccat 2500

Db   2548 AGTGGACCCCCTGGCCCAGGTGACTCAGCTATTCCGGGAACATCTCTTAGAGCGAGCACT 2607

Qy   2501 tcagtgtctggtaggagcgggccacaaatcgggaggcctgcccacatcttctgtcagcgg 2560

Db   2608 GAACTGTGTGACCCAGCCCAACCCCAGCCCTGGGTCAGCTGATGGGGACAAGGAATTCTC 2667

Qy   2561 agaggcggaacagttgcagcaacagcagcacagcggcaccattgtcagcaatgttcttaa 2620

Db   2668 GGATGCCCTCGGGTACCTGCAGCTGCTGAACAGCTGTTCTGATGCTGCGGGGGCTCCTGC 2727
```

```
Qy   2621 gtacacgtccctccttaaggacactctctgggctgatgaggatgagcgggatacaaacgt 2680
          ||||  |  |  | | |    ||    |  |||           |     ||  |    |
Db   2728 CTACAGCTTCTCCATCAGTTCCA---GCATGGCCACCACCACCGGCGTAGACCCGGTGGC 2784

Qy   2681 ggtgtggtgggccgatgttttggagaccgcagtgcactggctccttggtgaagacacgct 2740
            ||||||||||  | |    |     |   | ||||||| |   | || ||   |
Db   2785 CAAGTGGTGGGCCTCTCTGACAGCTGTGGTGATCCACTGGCTGCGGCGGGATGAGGAGGC 2844

Qy   2741 ggccgagcaattgtacggcaggatcaagcaaatgcccacgcagctgcaacagtgcggcga 2800
          ||| ||||   ||| |     | |  ||||  |||||  |  ||||||  |||  |
Db   2845 GGCTGAGCGGCTGTGCCCGCTGGTGGAGCACCTGCCCCGGGTGCTGCAGGAGTCTGAGAG 2904

Qy   2801 aaacgatcatctgcccaaggcgctgcatgctgtgctgcgagctaagatgatcttactaaa 2860
          |  |   |    | |    || |||||  |  |    |   ||   |    |  |
Db   2905 ACCCCTGCCCAGGGC---AGCTCTGCACTCCTTCAAGGCTGCCCGGGCCCTGCTGGGCTG 2961

Qy   2861 aaacaatggcaacgcactggacaaaagtctcaagcaattggtaaacatcctctgcgatga 2920
            ||| ||||  |    |  ||    |   | ||  ||  | |   ||  || |
Db   2962 TGCCAA-GGCAGAGTCTGGTCCAGCCAGCCTGACCATCTGTGAGAAGGCCAGTGGGTACC 3020

Qy   2921 gtcgagtgtggagctccaagagtgcttgactgtcaaccggatcaccgacgccaagggtat 2980
            |  |   |    | |  |   |    |   || |   ||   | | |||  |    |
Db   3021 TGCAGGACAGCCTGGCTACCACACCAG--CCAGCAGCTCCATTGACAAGGCCGTGCAGCT 3078

Qy   2981 aaagctgcttttccagttgcttacctgcgattggctgctcgaaactaggactgctctgtg 3040
            |||  |  || | | |||                ||          |      |
Db   3079 GTTCCTGTGTGACCTGCTTCTTGTGGTGCGCACCAGCCTGTGGCGGCAGCAGCAGCCCCC 3138

Qy   3041 ggaactggaacacatgaatatggaggacgatggcttctaccaagtgccaggtgaagtgct 3100
          ||  | ||  |           ||    |    ||     ||| |   | |     | |||
Db   3139 GGCCCCGGCCCCAGCAGCCCAGGGCACCAGCAGCAGGCCCCAGGCTTCCGCCCTTGAGCT 3198

Qy   3101 cgagaagttccagaccgatttgaactcgttgcgcaacattgtggagaatataccgaacgc 3160
               |||||    ||  ||| |   || |     | |   |||   | | |  |||
Db   3199 GCGTGGCTTCCAACGGGACCTGAGCAGCCTGAGGCGGCTGGCACAGAGCTTCCGGCCCGC 3258

Qy   3161 ccaatcgcgcatatatttgtacgaggcagtttgtcgcctgatggctggagcctcaccgtg 3220
          |     | |  | |   |  | |||||       || ||||||| || |||   ||
Db   3259 CATGCGGAGGGTGTTCCTACATGAGGCCACGGCCCGGCTGATGGCGGGGGCCAGCCCCAC 3318

Qy   3221 tccaacgcaacagctcttggacaggagtctgcgatcacgcaacgcccactcgtccatctt 3280
            |  || || |||||| | ||| | ||||||  |    ||    | || |     ||
Db   3319 ACGGACACACCAGCTCCTCGACCGCAGTCTGAGGCGGCGGGCAGGCCCCGGTGGCAAAGG 3378

Qy   3281 ctgcggcagcaaggatcggcggcagcagaacttcgtgggcggagagcgggaacgggcttc 3340
            |||       ||| | |  || ||          | |    ||||| |  ||
Db   3379 AGGCGCGGTGGCGGAGCTGGAGCCGC---------GGCCCACGCGGCGGGAGCACGCGGA 3429

Qy   3341 ggccatgtacgtggcctgcaagtatctcccgcctgcgctgctcagctccccgggtgaacg 3400
          |||| ||    |||||| |   || || || || |   | ||     | || ||  | ||
Db   3430 GGCCTTGCTGCTGGCCTCCTGCTACCTGCCCCCCGGCTTCCTGTCGGCGCCCGGGCAGCG 3489

Qy   3401 tgctggcatgttagccgaggcggccaagaccctggagaaggtgggcgacaagcgaaagct 3460
            |  |||||| | || |||||||||    || || |||||| | |||||    ||   |||
Db   3490 CGTGGGCATGCTGGCTGAGGCGGCGCGCACACTCGAGAAGCTTGGCGATCGCCGGCTGCT 3549

Qy   3461 caaggagtgctaccagctgatgaagtcgctgggcaacggca 3501
            | || ||  | ||| || | | |   ||||||   || |
Db   3550 GCACGACTGTCAGCAGATGCTCATGCGCCTGGGCGGTGGGA 3590


RESULT  11
HSU02031
LOCUS       HSU02031     4249 bp    mRNA            PRI       22-OCT-1994
DEFINITION  Human sterol regulatory element binding protein-2 mRNA,  complete
            cds.
ACCESSION   U02031
VERSION     U02031.1  GI:451329
KEYWORDS    SREBP-2.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1  (bases 1 to 4249)
  AUTHORS   Hua,X., Yokoyama,C., Wu,J., Briggs,M.R., Brown,M.S., Goldstein,J.L.
            and Wang,X.
  TITLE     SREBP-2, a second basic-helix-loop-helix-leucine zipper protein
            that stimulates transcription by binding to a sterol regulatory
            element
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 90 (24), 11603-11607 (1993)
  MEDLINE   94089681
REFERENCE   2  (bases 1 to 4249)
  AUTHORS   Hua,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1993) Xianxin Hua, Department of Molecular
            Genetics, University of Texas, Southwestern Medical Center, 5323
            Harry Hines Blvd., Dallas, TX 75235 USA
FEATURES             Location/Qualifiers
     source          1. .4249
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             118. .3543
                     /standard_name="SREBP-2"
                     /codon_start=1
                     /product="sterol regulatory element binding protein-2"
                     /protein_id="AAA50746.1"
                     /db_xref="GI:451330"
                     /translation="MDDSGELGGLETMETLTELGDELTLGDIDEMLQFVSNQVGEFPD
                     LFSEQLCSSFPGSGGSGSSSGSGSSSSSSSNGRGSSSGAVDPSVQRSFTQVTLPSFSP
                     SAASPQAPTLQVKVSPTSVPTTPRATPILQPRPQPQPQPQTQLQQQTVMITPTFSTTP
                     QTRIIQQPLIYQNAATSFQVLQPQVQSLVTSSQVQPVTIQQQVQTVQAQRVLTQTANG
                     TLQTLAPATVQTVAAPQVQQVPVLVQPQIIKTDSLVLTTLKTDGSPVMAAVQNPALTA
                     LTTPIQTAALQVPTLVGSSGTILTTMPVMMGQEKVPIKQVPGGVKQLEPPKEGERRTT
                     HNIIEKRYRSSINDKIIELKDLVMGTDAKMHKSGVLRKAIDYIKYLQQVNHKLRQENM
                     VLKLANQKNKLLKGIDLGSLVDNEVDLKIEDFNQNVLLMSPPASDSGSQAGFSPYSID
                     SEPGSPLLDDAKVKDEPDSPPVALGMVDRSRILLCVLTFLCLSFNPLTSLLQWGGAHD
                     SDQHPHSGSGRSVLSFESGSGGWFDWMMPTLLLWLVNGVIVLSVFVKLLVHGEPVIRP
                     HSRSSVTFWRHRKQADLDLARGDFAAAAANLQTCLAVLGRALPTSRLDLACSLSWNVI
                     RYSLQKLRLVRWLLKKVFQCRRATPATEAGFEDEAKTSARDAALAYHRLHQLHITGKL
                     PAGSACSDVHMALCAVNLAECAEEKIPPSTLVEIHLTAAMGLKTRCGGKLGFLASYFL
                     SRAQSLCGPEHSAVPDSLRWLCHPLGQKFFMERSWSVKSAAKESLYCAQRNPADPIAQ
                     VHQAFCKNLLERAIESLVKPQAKKKAGDQEEESCEFSSALEYLKLLHSFVDSVGVMSP
                     PLSRSSVLKSALGPDIICRWWTSAITVAISWLQGDDAAVRSHFTKVERIPKALEVTES
                     PLVKAIFHACRAMHASLPGKADGQQSSFCHCERASGHLWSSLNVSGGTSDPALNHVVQ
                     LLTCDLLLSLRTALWQKQASASQAVGETYHASGAELAGFQRDLGSLRRLAHSFRPAYR
                     KVFLHEATVRLMAGGSPTRTHQLLEHSLRRRTTQSTKHGEVDAWPGQRERATAILLAC
                     RHLPLSFLSSPGQRAVLLAEAARTLEKVGDRRSCNDCQQMIVKLGGGTAIAAS"
BASE COUNT      860 a   1315 c   1202 g    872 t
ORIGIN


  Query Match             3.0%;  Score 114.4;  DB 9;  Length 4249;
  Best Local Similarity   52.3%;  Pred. No. 1.7e-16;
  Matches  315;  Conservative    0;  Mismatches  266;  Indels   21;  Gaps    2;

Qy   2909 cctctgcgatgagtcgagtgtggagctccaagagtgcttgactgtcaaccggatcaccga 2968
          ||  |||||   | | ||||   | ||        || | |  ||||   ||  ||||
Db   2946 CCATTGCGAGAGGGCCAGTGGCCACCTATGGAGCAGCCTCAACGTCAGTGGGGGCACCTC 3005

Qy   2969 cgccaagggtataaagctgcttttccagttgcttacctgcgattggctgctcgaaactag 3028
            | |   |   | || |   |  ||||| |||| ||||| ||   ||| ||        |
Db   3006 TGACCCTGCCCTCAACCACGTGGTCCAGCTGCTCACCTGTGACCTGCTACTGTCGCTACG 3065

Qy   3029 gactgctctgtgggaactggaacacatgaatatggaggacgatgg---cttctaccaagt 3085
          ||| || || ||| ||    |   ||     |  |||  |  ||     ||||||  |
Db   3066 GACAGCGCTCTGGCAAAAACAGGCCAGTGCCAGCCAGGCTGTGGGGGAGACCTACCACGC 3125

Qy   3086 gccaggtgaagtgctcgagaagttccagaccgatttgaactcgttgcgcaacattgtgga 3145
          | |||| |  |  || | |   |||||    ||  ||  |    ||||||   | |   |
Db   3126 GTCAGGCGCTGAACTGGCGGGCTTCCAACGGGACCTGGGCAGCCTGCGCAGGCTGGCACA 3185

Qy   3146 gaatataccgaacgcccaatcgcgcatatatttgtacgaggcagtttgtcgcctgatggc 3205
            |   | |     ||  |       | |   || | || ||       ||||||||||
Db   3186 CAGCTTCCGCCCAGCATACCGCAAGGTGTTCCTGCATGAAGCCACCGTGCGCCTGATGGC 3245

Qy   3206 tggagcctcaccgtgtccaacgcaacagctcttggacaggagtctgcgatcacgcaacgc 3265
```

```
Db   3246 AGGAGGCAGCCCCACCCGCACCCACCAGCTGCTGGAACACAGCCTGCGGCGGCGCACCAC 3305

Qy   3266 ccactcgtccatcttctgcggcagcaaggatcggcggcagcagaacttcgtgggcggaga 3325
Db   3306 GCA-----------------GAGCACCAAGCACGGAGAGGTGGATGCCTGGCCCGGCCA 3347

Qy   3326 gcgggaacgggcttcggccatgtacgtggcctgcaagtatctcccgcctgcgctgctcag 3385
Db   3348 GCGAGAGCGGGCCACCGCCATCCTGCTGGCCTGCCGCCACCTGCCCCTCTCCTTCCTCTC 3407

Qy   3386 ctccccgggtgaacgtgctggcatgttagccgaggcggccaagaccctggagaaggtggg 3445
Db   3408 CTCCCCGGGCCAGCGGGCAGTGCTGCTGGCCGAAGCTGCCCGCACCCTGGAGAAGGTGGG 3467

Qy   3446 cgacaagcgaaagctcaaggagtgctaccagctgatgaagtcgctgggcaacggcattgg 3505
Db   3468 CGACCGGCGCTCCTGCAACGACTGCCAGCAGATGATTGTTAAGCTGGGTGGTGGCACTGC 3527

Qy   3506 ca 3507
Db   3528 CA 3529


RESULT 12
AR069307
LOCUS       AR069307     5197 bp    DNA             PAT       18-FEB-2000
DEFINITION  Sequence 53 from patent US 5891631.
ACCESSION   AR069307
VERSION     AR069307.1  GI:7220195
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5197)
  AUTHORS   Goldstein,J.L., Brown,M.S., Briggs,M.R. and Wang,X.
  TITLE     Methods relating tosterol regulatory element binding proteins
  JOURNAL   Patent: US 5891631-A 53 06-APR-1999;
FEATURES             Location/Qualifiers
     source          1. .5197
                     /organism="unknown"
BASE COUNT     1066 a   1570 c   1434 g   1127 t
ORIGIN


  Query Match             3.0%;  Score 114.4;  DB 5;  Length 5197;
  Best Local Similarity   52.3%;  Pred. No. 1.8e-16;
  Matches  315;  Conservative    0;  Mismatches  266;  Indels   21;  Gaps    2;

Qy   2909 cctctgcgatgagtcgagtgtggagctccaagagtgcttgactgtcaaccggatcaccga 2968
Db   2946 CCATTGCGAGAGGGCCAGTGGCCACCTATGGAGCAGCCTCAACGTCAGTGGGGGCACCTC 3005

Qy   2969 cgccaagggtataaagctgcttttccagttgcttacctgcgattggctgctcgaaactag 3028
Db   3006 TGACCCTGCCCTCAACCACGTGGTCCAGCTGCTCACCTGTGACCTGCTACTGTCGCTACG 3065

Qy   3029 gactgctctgtgggaactggaacacatgaatatggaggacgatgg---cttctaccaagt 3085
Db   3066 GACAGCGCTCTGGCAAAAACAGGCCAGTGCCAGCCAGGCTGTGGGGGAGACCTACCACGC 3125

Qy   3086 gccaggtgaagtgctcgagaagttccagaccgatttgaactcgttgcgcaacattgtgga 3145
Db   3126 GTCAGGCGCTGAACTGGCGGGCTTCCAACGGGACCTGGGCAGCCTGCGCAGGCTGGCACA 3185

Qy   3146 gaatataccgaacgcccaatcgcgcatatatttgtacgaggcagtttgtcgcctgatggc 3205
Db   3186 CAGCTTCCGCCCAGCATACCGCAAGGTGTTCCTGCATGAAGCCACCGTGCGCCTGATGGC 3245

Qy   3206 tggagcctcaccgtgtccaacgcaacagctcttggacaggagtctgcgatcacgcaacgc 3265
Db   3246 AGGAGGCAGCCCCACCCGCACCCACCAGCTGCTGGAACACAGCCTGCGGCGGCGCACCAC 3305

Qy   3266 ccactcgtccatcttctgcggcagcaaggatcggcggcagcagaacttcgtgggcggaga 3325
Db   3306 GCA-----------------GAGCACCAAGCACGGAGAGGTGGATGCCTGGCCCGGCCA 3347

Qy   3326 gcgggaacgggcttcggccatgtacgtggcctgcaagtatctcccgcctgcgctgctcag 3385
Db   3348 GCGAGAGCGGGCCACCGCCATCCTGCTGGCCTGCCGCCACCTGCCCCTCTCCTTCCTCTC 3407

Qy   3386 ctccccgggtgaacgtgctggcatgttagccgaggcggccaagaccctggagaaggtggg 3445
Db   3408 CTCCCCGGGCCAGCGGGCAGTGCTGCTGGCCGAAGCTGCCCGCACCCTGGAGAAGGTGGG 3467

Qy   3446 cgacaagcgaaagctcaaggagtgctaccagctgatgaagtcgctgggcaacggcattgg 3505
Db   3468 CGACCGGCGCTCCTGCAACGACTGCCAGCAGATGATTGTTAAGCTGGGTGGTGGCACTGC 3527

Qy   3506 ca 3507
Db   3528 CA 3529


RESULT 13
I22319
LOCUS       I22319       5197 bp    DNA             PAT       07-OCT-1996
DEFINITION  Sequence 53 from patent US 5527690.
ACCESSION   I22319
VERSION     I22319.1  GI:1602673
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5197)
  AUTHORS   Goldstein,J.L., Brown,M.S., Briggs,M.R. and Wang,X.
  TITLE     Methods and compositions relating to sterol regulatory element
            binding proteins
  JOURNAL   Patent: US 5527690-A 53 18-JUN-1996;
FEATURES             Location/Qualifiers
     source          1. .5197
                     /organism="unknown"
BASE COUNT     1066 a   1570 c   1434 g   1127 t
ORIGIN


  Query Match             3.0%;  Score 114.4;  DB 5;  Length 5197;
  Best Local Similarity   52.3%;  Pred. No. 1.8e-16;
  Matches  315;  Conservative    0;  Mismatches  266;  Indels   21;  Gaps    2;

Qy   2909 cctctgcgatgagtcgagtgtggagctccaagagtgcttgactgtcaaccggatcaccga 2968
Db   2946 CCATTGCGAGAGGGCCAGTGGCCACCTATGGAGCAGCCTCAACGTCAGTGGGGGCACCTC 3005

Qy   2969 cgccaagggtataaagctgcttttccagttgcttacctgcgattggctgctcgaaactag 3028
Db   3006 TGACCCTGCCCTCAACCACGTGGTCCAGCTGCTCACCTGTGACCTGCTACTGTCGCTACG 3065

Qy   3029 gactgctctgtgggaactggaacacatgaatatggaggacgatgg---cttctaccaagt 3085
Db   3066 GACAGCGCTCTGGCAAAAACAGGCCAGTGCCAGCCAGGCTGTGGGGGAGACCTACCACGC 3125

Qy   3086 gccaggtgaagtgctcgagaagttccagaccgatttgaactcgttgcgcaacattgtgga 3145
Db   3126 GTCAGGCGCTGAACTGGCGGGCTTCCAACGGGACCTGGGCAGCCTGCGCAGGCTGGCACA 3185

Qy   3146 gaatataccgaacgcccaatcgcgcatatatttgtacgaggcagtttgtcgcctgatggc 3205
Db   3186 CAGCTTCCGCCCAGCATACCGCAAGGTGTTCCTGCATGAAGCCACCGTGCGCCTGATGGC 3245

Qy   3206 tggagcctcaccgtgtccaacgcaacagctcttggacaggagtctgcgatcacgcaacgc 3265
Db   3246 AGGAGGCAGCCCCACCCGCACCCACCAGCTGCTGGAACACAGCCTGCGGCGGCGCACCAC 3305

Qy   3266 ccactcgtccatcttctgcggcagcaaggatcggcggcagcagaacttcgtgggcggaga 3325
```

```
Db    3306 GCA------------------GAGCACCAAGCACGGAGAGGTGGATGCCTGGCCCGGCCA 3347

Qy    3326 gcgggaacgggcttcggccatgtacgtggcctgcaagtatctcccgcctgcgctgctcag 3385
           ||| || |||||  | |||||     ||||||||    | || || |   |  | |||
Db    3348 GCGAGAGCGGGCCACCGCCATCCTGCTGGCCTGCCGCCACCTGCCCCTCTCCTTCCTCTC 3407

Qy    3386 ctccccgggtgaacgtgctggcatgttagccgaggcggccaagaccctggagaaggtggg 3445
           |||||||||  | || || |   || | ||||| || |||   ||||||||||||||||
Db    3408 CTCCCCGGGCCAGCGGGCAGTGCTGCTGGCCGAAGCTGCCCGCACCCTGGAGAAGGTGGG 3467

Qy    3446 cgacaagcgaaagctcaaggagtgctaccagctgatgaagtcgctgggcaacggcattgg 3505
           ||||  |||      ||| || ||| | ||| ||||      ||||||    |||| ||
Db    3468 CGACCGGCGCTCCTGCAACGACTGCCAGCAGATGATTGTTAAGCTGGGTGGTGGCACTGC 3527

Qy    3506 ca 3507
           ||
Db    3528 CA 3529


RESULT 14
CGU09103
LOCUS       CGU09103     3816 bp    mRNA            ROD       26-JUL-1994
DEFINITION  Cricetulus griseus sterol regulatory element binding protein-1
            (SREBP-1) mRNA, complete cds.
ACCESSION   U09103
VERSION     U09103.1  GI:516002
KEYWORDS    .
SOURCE      Chinese hamster.
  ORGANISM  Cricetulus griseus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
REFERENCE   1  (bases 1 to 3816)
  AUTHORS   Sato,R., Yang,J., Wang,X., Evans,M.J., Ho,Y.K., Goldstein,J.L. and
            Brown,M.S.
  TITLE     Assignment of the membrane attachment, DNA binding, and
            transcriptional activation domains of sterol regulatory binding
            protein-1 (SREBP-1)
  JOURNAL   J. Biol. Chem. 269, 17267-17273 (1994)
  MEDLINE   94274723
REFERENCE   2  (bases 1 to 3816)
  AUTHORS   Yang,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-1994) Jianxin Yang, Molecular Genetics,
            University of Texas Southwestern Medical Center, 5323 Harry Hines
            Blvd., Dallas, TX 75237-9046, USA
FEATURES             Location/Qualifiers
     source          1. .3816
                     /organism="Cricetulus griseus"
                     /db_xref="taxon:10029"
                     /cell_line="CHO"
                     /cell_type="CHO K1"
     CDS             31. .3432
                     /note="SREBP-1"
                     /codon_start=1
                     /product="sterol regulatory element binding protein-1"
                     /protein_id="AAA20085.1"
                     /db_xref="GI:516003"
                     /translation="MDELPFGEAAVEQALDELGELDAALLTDIQDMLQLINNQDSDFP
                     GLFDSPYAGGGAGDTEPTSPGANSPESLSSPASLGSSLEAFLGEPKATPASLSPVPSA
                     STALKMYPSVPPFSPGPGIKEEPVPLTILQPPAAQPSPGTLLPPSFPPPPLQLSPAPV
                     LGYSSLPSGFSGTLPGNTQQPPSSLSLASAPGVSPISLHTQVQSSASQQPLPASTAPR
                     TTTVTSQIQRVPVVLQPHFIKADSLLLTTVKTDTGATMKTAGISTLAPGTAVQAGPLQ
                     TLVSGGTILATVPLVVDTDKLPIHRLAAGSKALGSAQSRGEKRTAHNAIEKRYRSSIN
                     DKIVELKDLVVGTEAKLNKSAVLRKAIDYIRFLQHSNQKLKQENLALRNAAHKSKSLK
                     DLVSACGSAGGTDVAMEGVKPEVVDTLTPPPSDAGSPSQSSPLSLGSRGSSSGGSDSE
                     PDSPVFEDSQVKAQRLHSHGMLDRSRLALCALVFLCLTCNPLASLFGWGIPGPSSASG
                     AHHSSGRSMLEAESRDGSNWTQWLLPPLVWLANGLLVLACLALLFVYGEPVTRPHTSP
                     AVHFWRHRKQADLDLARGDFAQAAQQLWLALQALGRPLPTSNLDLACSLLWNLIRHLL
                     QRLWVGRWLAGRAGGLRRDCGLRMDARASARDAALVYHKLHQLHAMGKYTGGHLIASN
                     LALSALNLAECAGDAVSMATLAEIYVAAALRVKTSLPRALHFLTRFFLSSARQACLAQ
                     SGSVPLAMQWLCHPVGHRFFVDGDWAVHGAPQESLYSVAGNPVDPLAQVTRLFCEHLL
                     ERALNCIAQPSPGTADGDREFSDALGYLQLLNRCSDAVGTPACSFSVSSSMASTTGTD
                     PVAKWWASLTAVVIHWLRRDEEAAERLYPLVERMPHVLQETERPLPKAALYSFKAARA
                     LLDHRKVESGPASLAICEKASGYLRDSLAAPPTGSSIDKAMQLLLCDLLLVARTSMWQ
                     RQQSPASAQVAHSASNGSQASALELRGFQQDLSSLRRLAQNFRPAMRRVFLHEATARL
                     MAGASPARTHQLLDRSLRRRAGSSGKGGTVAELEPRPTWREHTEALLLASCYLPPAFL
                     SAPGQQMSMLAEAARTVEKLGDHRLLLDCQQMLLRLGGGTTVTSS"
     polyA_signal    3793. .3799
     polyA_site      3816
                     /note="13 A residues"
BASE COUNT      770 a   1188 c   1061 g    797 t
ORIGIN


  Query Match             3.0%;  Score 112;  DB 12;  Length 3816;
  Best Local Similarity   44.1%;  Pred. No. 6.4e-16;
  Matches 1096;  Conservative    0;  Mismatches 1320;  Indels   72;  Gaps   12;

Qy    1029 aagcgctcggcccacaacgccatcgagcggcgctatcgcacctcaatcaacgacaagatt 1088
           |||||| | |||||||| ||||| ||| ||||||  ||  |||| ||||| |||||||||
Db     982 AAGCGCACAGCCCACAATGCCATTGAGAAGCGCTACCGTTCCTCTATCAATGACAAGATT 1041

Qy    1089 aacgagttgaagaacttggtagtgggagagcaggccaagctgaacaagtccgcagtgttg 1148
             ||| |  ||  || |||| |||||     |||| |||||||| || ||  || || |||
Db    1042 GTGGAGCTCAAAGACCTGGTGGTGGGCACTGAGGCAAAGCTGAATAAATCTGCCGTCTTG 1101

Qy    1149 cggaaatccatagacaagattcgggatctgcaacgccagaatcacgatctgaaggcagag 1208
           || ||  |||| ||| | || ||     | || | |   || ||  | || |||   |||
Db    1102 CGCAAGGCCATCGACTATATCCGCTTCTTACAGCACAGCAACCAGAAGCTCAAGCAGGAG 1161

Qy    1209 ttgcagcgcctgcagagggagctaatggcacgcgacggctccaaggtgaaggatttactt 1268
             | |  |||          || ||  ||      ||     |||      | |    ||
Db    1162 AACCTGGCCCT---------GCGAAATGCCGCTCACAAAAGCAAATCCCTGAAGGACCTG 1212

Qy    1269 cagctgggcactcggcctggtagagcatccaagaagcgccgcgagagctcgcagaccttt 1328
             |  || |  | |   ||   |||   |  |   |       ||| |   | || |
Db    1213 GTGTCGGCCTGTGGCAGTGCAGGAGGCACAGATGTGGCTATGGAGGGTGTGAAGCCTGAG 1272

Qy    1329 accacggatgccggactgacgccgccacgcagcgatgaatcggatccttcgctctcgccc 1388
             |||| |        | || ||      || ||  |||     |   |       |||
Db    1273 GTGGTGGATACGCTGACCCCTCCACCCTCAGACGCTGGCTCGCCCTCCCAGAGTAGCCCC 1332

Qy    1389 atgcactcggacatctcgttgccgccatcaccctatggtggatccaccgccagctgtagc 1448
           ||  |   | || |  |  |  ||        |      ||  |   ||| |      |
Db    1333 TTGTCCCTCGGCAGC-AGAGGTAGCAGCAGTGGTGGCAGTGACTCGGAGCCTGACAGCCC 1391

Qy    1449 agtggcagcagcagcaatgaagaaccactggtggtgcccagctctatgcgcggcatg 1508
           |||    |  |         ||||  |||  ||  |  |        |  |  ||||||
Db    1392 AGTCTTTGAGGATAGCCAGGTGAAAGCCCAACGGCTGCAC-------AGTCATGGCATG 1443

Qy    1509 gccacccactctcgcctcggactctgcatgtttatgttcgccatcctggccgtcaatccc 1568
             || ||| ||||| |  || ||   | |  | |||      ||| ||  ||| |||
Db    1444 CTGGACCGCTCCCGCCTAGCCCTGTGTGCGCTGGTCTTCCTGTGTCTGACCTGCAACCCC 1503

Qy    1569 ttcaagac---ctttctccagcgcggccactatgacagtaatgacgatcttggcgacatg 1625
           ||     |   ||| |  | ||  || |  |  |  | || |  |  ||   ||
Db    1504 TTGGCATCACTGTTTGGCTGGGGCATCCCCGGTCCCTCCAGTGCCTCTGGTGCACACCAC 1563

Qy    1626 agcggtcaaagacgcattctctcttacga---cgtggaaggtgaaggttttgctgtctgg 1682
           |||  |    |  |||| ||      |||   |   || ||      ||  |   |||
Db    1564 AGCTCTGGGCGTAGCATGCTGGAGGCCGAGAGCAGAGATGGCTCTAATTGGACCCAGTGG 1623

Qy    1683 cagcagagttcctggatatggctattgaacttcacactgatgcttggatgcttggtgaaa 1742
             || |    ||    | |||||    ||      |||  || | |  ||| |||
Db    1624 TTGCTGCCACCCCTAGTCTGGCTGGCCAATGGACTACTAGTGTTGGCCTGCCTGGCTCTT 1683

Qy    1743 ttgctggtttacggtgatccgcagctggacgcgcaaacggacgc---------ctactgc 1793
           |  | || || || ||  |       | || ||   | |             || |||
Db    1684 CTCTTTGTCTATGGGGAACCTGTGACCCGGCCACACACTAGCCCAGCTGTACACTTCTGG 1743

Qy    1794 cagcacaggcagcgggctgacttctattttagccaaggacagtcgtctcaggcctacgcc 1853
             ||  |  | | ||||||| |  | ||    |  ||| | |   | |||||
Db    1744 AGACATCGCAAACAGGCTGACCTGGACTTGGCTCGGGGAGATTTTGCCCAGGCTGCTCAG 1803

Qy    1854 ggttacctcaactgtctgcatatgtttggattaagtctaccggcgtcgcgcttggagtgt 1913
                       |||||    || |||      || ||  | |||  | | || |
```

```
Db   1804 CAGCTGTGGCTGGCCCTGCAGGCATTGGGACGGCCCCTGCCCACCTCGAACCTAGACTTG 1863

Qy   1914 tacttgcagaccacgtggcagttccttcgttttcttttccatcgcctctggctgggtcgg 1973
Db   1864 GCCTGCAGCCTGCTTTGGAACCTCATCCGCCACCTGCTGCAGCGTCTCTGGGTTGGCCGC 1923

Qy   1974 gtgctgtcacggcggtccggtgggctgtttagcaacgccgccagcaggaaacaggcgctg 2033
Db   1924 TGGCTGGCAGGCCGGGCTGGGGGCTTGCGGAGAGACTGTGGACTGAGAATGGATGCACGT 1983

Qy   2034 gcatctgcacgcgaactggccctgctcttcaaccgactgaatcaattgcaactgactgga 2093
Db   1984 GCCAGTGCTCGAGATGCGGCTCTCGTCTACCATAAGCTGCACCAGCTGCATGCCATGGGC 2043

Qy   2094 aatggaagccgcggtgacatgaacggcattatgatggcactattcgcaagcaacatggct 2153
Db   2044 AAATACACAGGAGGGCACCTCATTGCTTCTAACCTGGCACTGAGTGCCCTGAACCTGGCC 2103

Qy   2154 gaagtggcgcacaatctactgacaccgcgcgagaccatctgcatccacgtaatgacagcg 2213
Db   2104 GAGTGCGCAGGAGATGCTGTATCCATGGCAACGCTGGCAGAGATCTATGTGGCTGCTGCC 2163

Qy   2214 ttgcgaatgaagcgcagtgccccaaaatggttgcaacagttcttcgcccgatactacatg 2273
Db   2164 CTGAGGGTCAAGACCAGTCTCCCAAGAGCCTTG---CACTTTTTGACACGTTTCTTCCTG 2220

Qy   2274 agccgggctcgtcaagagtgcggtcgcactagggccaccgagcaaacgcaggagctacgt 2333
Db   2221 AGTAGTGCCCGCCAGGCCTGCCTGGCACAGAGTGGCTCAGTGC---CTCTTGCCATGCAG 2277

Qy   2334 tgggcattcacagcctatggatatcgctactgcgccacgcacgtcttcacgtacgatctg 2393
Db   2278 TGGCTCTGCCACCCTGTAGGCCACCGTTTCTTCGTGGATGGGGACTGGGCTGTGCATG-- 2335

Qy   2394 agcgactccggcgagcaggatggattcttcacacgtcttaggaatccatgtgatcccgct 2453
Db   2336 -------GTGCCCCACAGGAGAGCCTGTACAGCGTGGCTGGGAACCCAGTGGATCCCCTC 2388

Qy   2454 gcccacgtcattaagcaatatcgagagcatttgctgtttaaatccattcagtgtctggta 2513
Db   2389 GCCCAGGTGACTCGACTATTCTGCGAACATCTCTTGGAGAGAGCACTGAACTGTATTGCT 2448

Qy   2514 ggagcgggccacaaatcgggaggcctgcccacatcttctgtcagcggagaggcggaacag 2573
Db   2449 CAACCCAGCCCGGGGACAGCTGATGGAGACAGGGAGTTCTCTGACGCACTTGGATACCTG 2508

Qy   2574 ttgcagcaacagcagcacagcggcaccattgtcagcaatgttcttaagtacacgtccctc 2633
Db   2509 CAGTTGCTAAATCGCTGCTCTGATGCTGTCGGGACTCCTGCCTGCAGCTTCTCTGTCAGC 2568

Qy   2634 cttaaggacactctctgggctgatgaggatgagcgggatacaaacgtggtgtggtgggcc 2693
Db   2569 TCCAG---------CATGGCTTCCACCACCGGCACAGACCCAGTGGCCAAGTGGTGGGCC 2619

Qy   2694 gatgttttggagaccgcagtgcactggctccttggtgaagacacgctggccgagcaattg 2753
Db   2620 TCACTGACGGCTGTGGTGATCCACTGGCTGCGGCGGGATGAAGAGGCAGCTGAGCGCCTA 2679

Qy   2754 tacggcaggatcaagcaaatgcccacgcagctgcaacagtgcggcgaaaacgatcatctg 2813
Db   2680 TACCCGCTGGTAGAGCGTATGCCCCACGTGCTGCAGGAGACTGAGAGACCC------CTG 2733

Qy   2814 cccaaggcgctgcatgctgtgctgcgagctaagatgatcttactaaaaaacaatggcaac 2873
Db   2734 CCCAAGGCAGCTCTGTACTCCTTCAAGGCTGCCCGGGCTCTGCTGGACCACAGAAAAGTG 2793

Qy   2874 gcactggacaaaagtctcaagcaattggtaaacatcctctgcgatgagtcgagtgtggag 2933
Db   2794 GAGTCTGGCCCAGCCAGCCTGGCCATCTGTGAGAAGGCCAGCGGGTACTTGCGGGACAGC 2853

Qy   2934 ctccaagagtgcttgactgtcaaccggatcaccgacgccaagggtataaagctgcttttc 2993
Db   2854 TTAGCCGCTCCACCAACTGGCAGCTCCATTGACAAGGCCATGCAGCTGCTCCTGTGTGAT 2913

Qy   2994 cagttgcttacctgcgattggctgctcgaaactaggactgctctgtgggaactggaacac 3053
Db   2914 CTACTTCTTGTGGCCCGCACTAGTATGTGGCAGCGCCAGCAGTCACCAGCCTCAGCCCAG 2973

Qy   3054 atgaatatggaggacgatggcttctaccaagtgccaggtgaagtgctcgagaagttccag 3113
Db   2974 GTAGCTCACAGTGCCAGCAATGGATCTCAGGCCTCCGCTTTGGAGCTTCGAGGTTTCCAA 3033

Qy   3114 accgatttgaactcgttgcgcaacattgtggagaatataccgaacgcccaatcgcgcata 3173
Db   3034 CAGGACCTGAGCAGCCTGAGGCGCTTGGCACAGAACTTCCGGCCTGCTATGAGGAGAGTG 3093

Qy   3174 tatttgtacgaggcagtttgtcgcctgatggctggagcctcaccgtgtccaacgcaacag 3233
Db   3094 TTCCTACACGAGGCCACAGCTCGGCTGATGGCAGGGGCAAGTCCTGCCCGGACACACCAG 3153

Qy   3234 ctcttggacaggagtctgcgatcacgcaacgcccactcgtccatcttctgcggcagcaag 3293
Db   3154 CTCCTGGACCGAAGTCTGCGGAGGCGGGCCGGCTCCAGTGGCAAAGGAGGCACTGTAGCT 3213

Qy   3294 gatcggcggcagcagaacttcgtgggcggagagcgggaacgggcttcggccatgtacgtg 3353
Db   3214 GAGCTGGAGCCTC---------GACCCACATGGCGGGAGCACACAGAGGCCTTGCTGCTG 3264

Qy   3354 gcctgcaagtatctcccgcctgcgctgctcagctccccgggtgaacgtgctggcatgtta 3413
Db   3265 GCCTCCTGCTATCTGCCACCTGCCTTCCTGTCGGCCCCTGGACAGCAAATGAGCATGTTG 3324

Qy   3414 gccgaggcggccaagaccctggagaaggtgggcgacaagcgaaagctcaaggagtgctac 3473
Db   3325 GCTGAGGCAGCACGCACTGTAGAGAAGCTTGGTGATCATCGGCTACTGCTTGACTGCCAG 3384

Qy   3474 cagctgatgaagtcgctgggcaacggca 3501
Db   3385 CAGATGCTTCTGCGCCTGGGCGGTGGGA 3412


RESULT  15
RATADD1A
LOCUS       RATADD1A     3115 bp    mRNA            ROD       08-SEP-1993
DEFINITION  Rat add1 mRNA sequence.
ACCESSION   L16995
VERSION     L16995.1  GI:294509
KEYWORDS    adipocyte protein; cell determination; cell differentiation;
            helix-loop-helix protein.
SOURCE      Rattus sp. cDNA to mRNA.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 3115)
  AUTHORS   Kim,J., Tontonoz,P. and Spiegelman,B.
  TITLE     Add1: a novel helix-loop-helix protein associated with adipocyte
            determination and differentiation
  JOURNAL   Mol. Cell. Biol. 13, 4753-4759 (1993)
  MEDLINE   93330269
FEATURES             Location/Qualifiers
     source          1. .3115
                     /organism="Rattus sp."
                     /db_xref="taxon:10118"
BASE COUNT      629 a   1002 c    870 g    614 t
ORIGIN


  Query Match             2.8%;  Score 105.8;  DB 12;  Length 3115;
  Best Local Similarity   45.7%;  Pred. No. 1.9e-14;
  Matches  586;  Conservative    0;  Mismatches  662;  Indels   33;  Gaps    5;

Qy   1029 aagcgctcggcccacaacgccatcgagcggcgctatcgcacctcaatcaacgacaagatt 1088
Db    925 AAGCGCACAGCCCACAATGCCATTGAGAAGCGCTACCGTTCCTCTATCAATGACAAGATT 984

Qy   1089 aacgagttgaagaacttggtagtgggagagcaggccaagctgaacaagtccgcagtgttg 1148
```

```
Db    985 GTGGAGCTCAAGGACCTGGTGGTGGGCACTGAGGCAAAGCTGAATAAATCTGCTGTCTTG 1044

Qy   1149 cggaaatccatagacaagattcgggatctgcaacgccagaatcacgatctgaaggcagag 1208

Db   1045 CGCAAGGCCATCGACTACATCCGCTTCTTACAGCACAGCAACCAGAAACTCAAGCAGGAG 1104

Qy   1209 ttgcagcgcctgcagagggagctaatggcacgcgacggctccaaggtgaaggatttactt 1268

Db   1105 AACCTGACCCT---------GCGAAGTGCTCACAAAAGCAAATCACTGAAAGA------C 1149

Qy   1269 cagctgggcactcggcctggtagagcatccaagaagcgccgcgagagctcgcagacctttt 1328

Db   1150 CTGGTGTCAGCTTGTGGCAGTGGAGGAGGCACAGATGTGTCTATGGAGGGCATGAAACCT 1209

Qy   1329 accacggatgccggactgacgccgccacgcagcgatgaatcggatccttcgctctcgccc 1388

Db   1210 GAAGTGGTAGAAACGCTGACCCCTCCACCCTCAGACGCCGGCTCACCCTCCCAGAGTAGC 1269

Qy   1389 atgcactcggacatctcgttgccgccatcaccctatggtggatccaccgccagctgtagc 1448

Db   1270 CCCTTGTCCTTGGGCAGCAGAGGCAGCAGCAGTGGTGGCAGTGACTCTGAGCCCGACAGC 1329

Qy   1449 agtggcagcagcagcagcaatgaagaaccactggtggtgcccagctctatgcgcggcatg 1508

Db   1330 CCAGCCTTTGAGGATAACCAGGTGAAAGCCCAGCGGCTGCCTTCACATAGCCGAGGCATG 1389

Qy   1509 gccacccactctcgcctcggactctgcatgtttatgttcgccatcctggccgtcaatccc 1568

Db   1390 CTGGACCCGTCCCGCCTGGCCCTGTGTGTACTGGTCTTCCTGTGTCTGACCTGCAACCCA 1449

Qy   1569 ttcaagacctttctccagcgcggccactatgacagtaatgacgatcttggcgacatgagc 1628

Db   1450 TTGGCCTCACTGTTTGGCTGGGGCATCCTCACTCCCTCTGATGCTTCGGGTGTGCACCGT 1509

Qy   1629 ggtcaaagacgcattctctcttacgacgtggaaggtgaagg------ttttgctgtctgg 1682

Db   1510 AGTTCTGGGCGCAGCATGCTGGAGGCCGAGAGCAGAGATGGCTCTAATTGGACCCAGTGG 1569

Qy   1683 cagcagagttcctggatatggctattgaacttcacactgatgcttggatgcttggtgaaa 1742

Db   1570 TTGCTGCCACCCCTAGTCTGGCTGGCCAATGGACTACTAGTGTTGGCCTGCTTGGCTCTT 1629

Qy   1743 ttgctggtttacggtgatcc---------gcagctggacgcgcaaacggacgcctactgc 1793

Db   1630 CTCTTTGTCTACGGGGAACCTGTGACCAGGCCACACTCCGGCCCGGCTGTACACTTCTGG 1689

Qy   1794 cagcacaggcagcgggctgacttctattttagccaaggacagtcgtctcaggcctacgcc 1853

Db   1690 AGACATCGCAAACAAGCTGACCTGGATTTGGCCCGGGGAGATTTTGCCCAGGCCGCTCAA 1749

Qy   1854 ggttacctcaactgtctgcatatgtttggattaagtctaccggcgtcgcgcttggagtgt 1913

Db   1750 CAGCTGTGGCTGGCCTTGCAAGCCCTGGGCCGGCCCCTGCCCACCTCAAACCTGGATCTG 1809

Qy   1914 tacttgcagaccacgtggcagttccttcgttttcttttccatcgcctctggctgggtcgg 1973

Db   1810 GCCTGCAGCCTGCTTTGGAACCTCGTCCGCCACCTGCTGCAGCGTCTTTGGGTGGGCCGC 1869

Qy   1974 gtgctgtcacggcggtccggtgggctgtttagcaacgccgccagcaggaaacaggcgctg 2033

Db   1870 TGGCTGGCAGGCCAGGCTGGGGGCCTGCAGAGGGACTACAGGCTGAGAAAGGATGCTCGT 1929

Qy   2034 gcatctgcacgcgaactggccctgctcttcaaccgactgaatcaattgcaactgactgga 2093

Db   1930 GCCAGTGCCCGAGATGCGGCTGTCGTCTACCATAAGCTGCACCAGCTGCATGCCATGGGC 1989

Qy   2094 aatggaagccgcggtgacatgaacggcattatgatggcactattcgcaagcaacatggct 2153

Db   1990 AAGTACACAGGAGGCCATCTTGTTGCTTCTAACCTGGCACTGAGTGCCCTTAACCTGGCT 2049

Qy   2154 gaagtggcgcacaatctactgacaccgcgcgagaccatctgcatccacgtaatgacagcg 2213

Db   2050 GAGTGTGCAGGAGATGCTATATCCATGGCAACACTGGCAGAGATCTACGTGGCAGCTGCC 2109

Qy   2214 ttgcgaatgaagcgcagtgccccaaaatggttgcaacagttcttcgcccgatactacatg 2273

Db   2110 CTAAGGGTCAAAACCAGCCTCCCCAGAGCCTTGCAC---TTCTTGACACGTTTCTTCCTA 2166

Qy   2274 agccgggctcgtcaagagtgc 2294

Db   2167 AGTAGTGCCCGCCAGGCCTGC 2187
```

Search completed: September 4, 2000, 01:11:57
Job time: 23737 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2000, 00:58:07 ; Search time 299.13 Seconds
(without alignments)
3151.552 Million cell updates/sec

Title: US-09-332-522B-7
Perfect score: 3768
Sequence: 1 gtttattaagctgcaaatat.....aatattttaacagatttaaa 3768

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

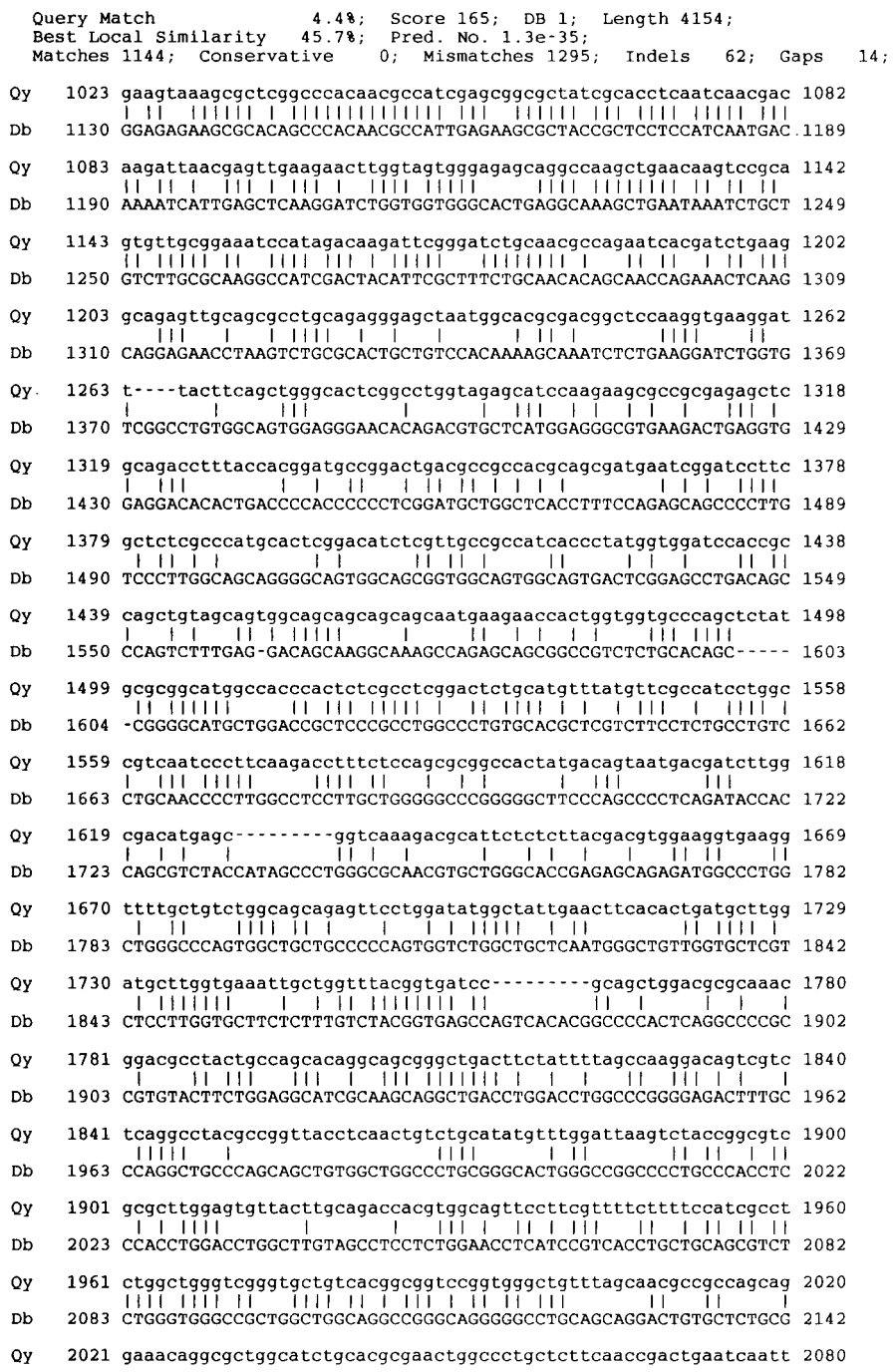
SUMMARIES

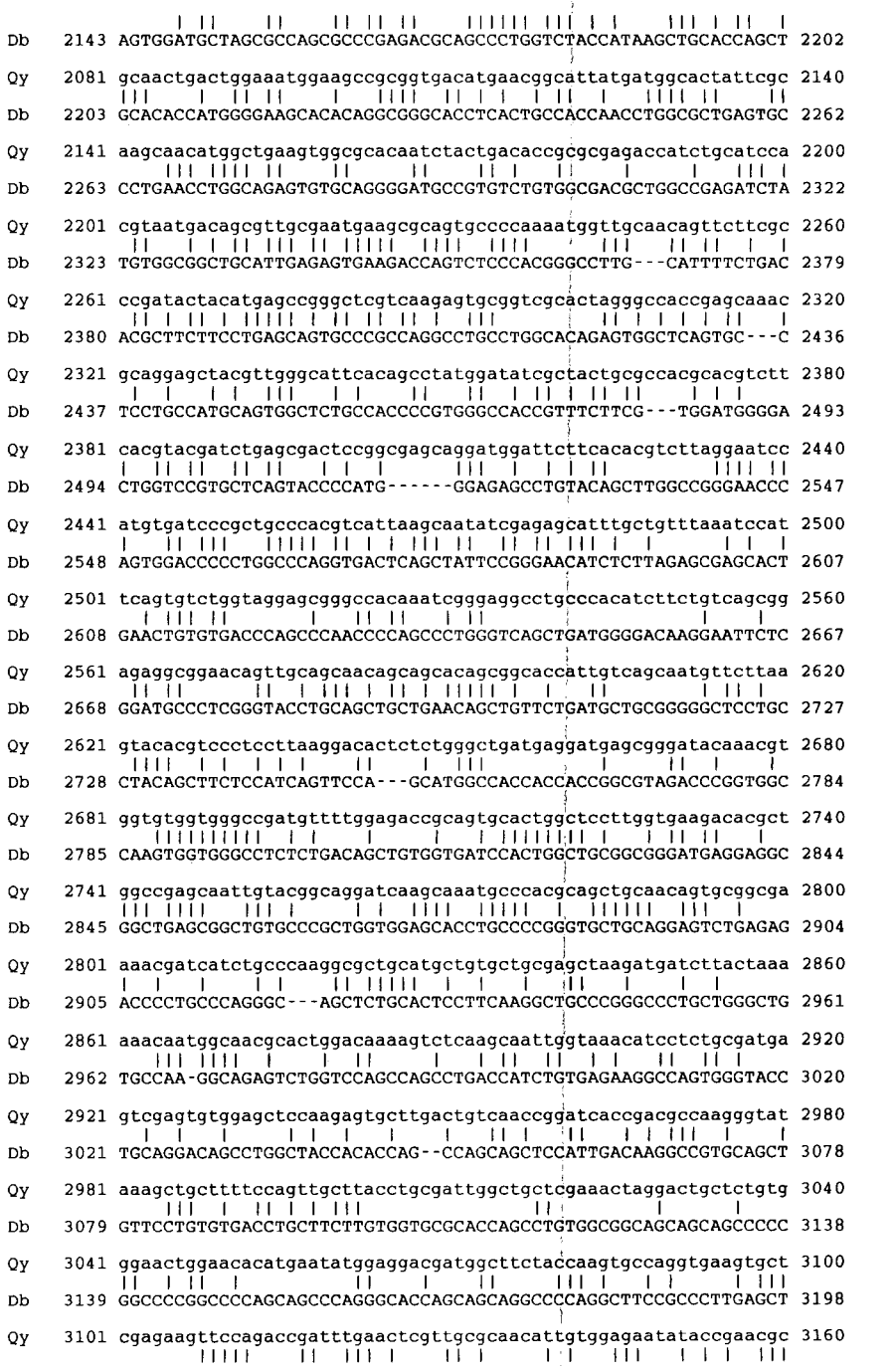
	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	165	4.4	4154	1	Q79037	Human SREBP-1a cod
	2	114.4	3.0	5197	1	Q79038	Human SREBP-2 codi
	3	48.4	1.3	234	1	Q84832	Spinocerebellar at
c	4	47.2	1.3	2744	1	Q98470	MiSP1-containing p
	5	47	1.2	203	1	V17226	SCA2 gene CAG repe
	6	47	1.2	203	1	V30271	Glutamine rich reg
	7	45.8	1.2	195	1	Q84831	Spinocerebellar at
c	8	44.8	1.2	32207	1	V73805	KSHV LUR DNA (nucl
c	9	44.8	1.2	137507	1	V19941	KSHV long unique c
	10	42.2	1.1	543	1	Q23092	Antigen tc-7a gene
	11	41.8	1.1	165	1	V30274	Glutamine rich reg
	12	41.2	1.1	14704	1	Q20685	PKS 741 insert con
	13	41	1.1	171	1	Q84834	Spinocerebellar at
	14	41	1.1	1498	1	X23749	Rice anthranilate
	15	40.6	1.1	154	1	Q84835	Spinocerebellar at
	16	40.6	1.1	168	1	Q84833	Spinocerebellar at
c	17	40	1.1	110000	1	X20248_00	Borrelia burgdorfe
c	18	39.8	1.1	1542	1	X16151	Human Sox1 encodin
	19	39.8	1.1	3879	1	V21060	cDNA encoding the
	20	39.8	1.1	3879	1	V29009	Human polyhomeotic
	21	39.8	1.1	10266	1	T33007	Mouse SRY-related
	22	39.2	1.0	726	1	X39871	Gastric cancer ass
c	23	38.8	1.0	786	1	T46148	T. cruzi L19E homo
c	24	38.2	1.0	50341	1	V22674	DNA sequence of a
c	25	38.2	1.0	52297	1	T51411	Mycobacteriophage
c	26	38.2	1.0	52298	1	Q47357	L5 mycobacteriopha
	27	38	1.0	506	1	Q84804	Spinocerebellar at
	28	38	1.0	3366	1	Q84803	Spinocerebellar at
	29	38	1.0	4126	1	X13277	Enterococcus faeca
	30	38	1.0	10660	1	Q84793	Spinocerebellar at
	31	37.4	1.0	2462	1	V35473	Human hSK3 coding
	32	37.2	1.0	1538	1	Q21732	DNA encoding a RAP
c	33	37.2	1.0	3211	1	V05541	Actinobacillus ple
	34	37.2	1.0	3284	1	V31337	Mortierella alpina
	35	36.8	1.0	245	1	T75504	P. americanus skin
	36	36.8	1.0	4356	1	Q37543	Cardiac adenylyl c
	37	36.8	1.0	4356	1	Q95540	Cardiac adenylyl c
c	38	36.8	1.0	114955	1	X53491	Human adenosine A1
	39	36.6	1.0	228	1	Q23094	Antigen tc-10a gen
	40	36.6	1.0	1239	1	Q61445	Lignin peroxidase
	41	36.6	1.0	1666	1	Q61443	Lignin peroxidase
	42	36.6	1.0	1810	1	Q51010	Lignin peroxidase
	43	36.6	1.0	1810	1	Q61444	Lignin peroxidase
	44	36.6	1.0	2051	1	X28069	Heparin sulphate 6
	45	36.6	1.0	2287	1	Q92617	Human afamin cDNA.

ALIGNMENTS

RESULT 1
Q79037
ID Q79037 standard; DNA; 4154 BP.
AC Q79037;
DT 04-AUG-1995 (first entry)
DE Human SREBP-1a coding sequence.
KW Probe; sterol regulatory element binding protein; regulatory protein;
KW cholesterol metabolism; sterol regulatory element-1; SREBP; SRE-1;
KW SREBP-1; SREBP-2; basic-helix-loop-helix-leucine zipper; plasma;
KW transcription factor; low density lipoporotein; LDL; receptor;
KW 3-hydroxy-3-methylglutaryl coenzyme A synthase; cholesterol;
KW hypercholesterolaemia; enhancer; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 167. .3610
FT /*tag= a
FT /product= SREBP-1a
PN WO9426922-A.
PD 24-NOV-1994.
PF 13-MAY-1994; U05300.
PR 13-MAY-1993; US-061697.
PR 01-OCT-1993; US-131365.
PA (TEXA) UNIV TEXAS SYSTEM.
PI Briggs MR, Brown MS, Goldstein JL, Wang X;
DR WPI; 95-006813/01.
DR P-PSDB; R66390.
PT New sterol regulator element binding protein - used to develop
PT prods. and screening assays for agents for reducing plasma
PT cholesterol levels (Eng)
PS Claim 31; Page 185-98; 305pp; English.
CC The sequences given in Q79037-38 encode the sterol regulatory element
CC binding proteins (SREBP), SREBP-1a and SREBP-2. SREBP's are regulatory
CC proteins which are involved in the regulation of genes involved in
CC cholesterol metabolism that are under the control of an associated
CC sterol regulatory element-1 (SRE-1) enhancer sequence. SREBP proteins
CC fall into two families, SREBP-1 and SREBP-2. Both proteins are members
CC of a family of basic-helix-loop-helix-leucine zipper (bHLH-Zip)
CC transcription factors. Each have the ability to bind to SRE sequences
CC and modulate SRE-mediated transcription. SRE-1 is a conditional enhancer
CC found in the promoters for the low density lipoporotein (LDL) receptor
CC and 3-hydroxy-3-methylglutaryl coenzyme A synthase genes. It increases
CC transcription in the absence of sterols and is inactivated when sterols
CC accumulate. Human SREBP-2 contins 1141 amino acids and has 47% identity
CC with STREBP-1a. SREBP-1a was the first recognised member of this family
CC and has 1147 amino acids. The resemblence between SREBP-1a and SREBP-2
CC includes an acidic N-terminus, a highly conserved bHLH-Zip motif (71%
CC identical), and an unusually long extension of 740 amino acids on the
CC C-terminal side of the bHLH-Zip region. SREBP-2 posseses one feature
CC lacking in SREBP-1a, a Glu rich region (27% Glu over 121 residues).
CC SREBP promotes SRE-1-mediated gene transcription, eg. LDL receptor
CC production in the presence of sterols. SREBP identified in screening
CC assays, may be used to reduce plasma cholesterol levels and in
CC controlling hypercholesterolaemia and its associated diseases.
SQ Sequence 4154 BP; 714 A; 1406 C; 1280 G; 754 T;

Query Match 4.4%; Score 165; DB 1; Length 4154;
Best Local Similarity 45.7%; Pred. No. 1.3e-35;
Matches 1144; Conservative 0; Mismatches 1295; Indels 62; Gaps 14;





```
Db    3199 GCGTGGCTTCCAACGGGACCTGAGCAGCCTGAGGCGGCTGGCACAGAGCTTCCGGCCCGC 3258

Qy    3161 ccaatcgcgcatatatttgtacgaggcagtttgtcgcctgatggctggagcctcaccgtg 3220
           |     | |  | |   |  | ||||        || |||||||| || |||   ||
Db    3259 CATGCGGAGGGTGTTCCTACATGAGGCCACGGCCCGGCTGATGGCGGGGGCCAGCCCCAC 3318

Qy    3221 tccaacgcaacagctcttggacaggagtctgcgatcacgcaacgcccactcgtccatctt 3280
           |  || || |||||| | ||| | |||||| |    ||    | || |     ||
Db    3319 ACGGACACACCAGCTCCTCGACCGCAGTCTGAGGCGGCGGGCAGGCCCCGGTGGCAAAGG 3378

Qy    3281 ctgcggcagcaaggatcggcggcagcagaacttcgtgggcggagagcgggaacgggcttc 3340
             |||       ||| | |  || ||          | |     |||||| |  ||
Db    3379 AGGCGCGGTGGCGGAGCTGGAGCCGC---------GGCCCACGCGGCGGGAGCACGCGGA 3429

Qy    3341 ggccatgtacgtggcctgcaagtatctcccgcctgcgctgctcagctccccgggtgaacg 3400
           |||| ||    ||||||| |   || || || || |   | ||     | || ||  | ||
Db    3430 GGCCTTGCTGCTGGCCTCCTGCTACCTGCCCCCCGGCTTCCTGTCGGCGCCCGGGCAGCG 3489

Qy    3401 tgctggcatgttagccgaggcggccaagaccctggagaaggtgggcgacaagcgaaagct 3460
           |  |||||| | || |||||||    || || |||||| | |||||    ||   |||
Db    3490 CGTGGGCATGCTGGCTGAGGCGGCGCGCACACTCGAGAAGCTTGGCGATCGCCGGCTGCT 3549

Qy    3461 caaggagtgctaccagctgatgaagtcgctgggcaacggca 3501
             | || ||  | ||| || | | |   ||||||   || |
Db    3550 GCACGACTGTCAGCAGATGCTCATGCGCCTGGGCGGTGGGA 3590


RESULT   2
Q79038
ID   Q79038 standard; DNA; 5197 BP.
AC   Q79038;
DT   04-AUG-1995  (first entry)
DE   Human SREBP-2 coding sequence.
KW   Probe; sterol regulatory element binding protein; regulatory protein;
KW   cholesterol metabolism; sterol regulatory element-1; SREBP; SRE-1;
KW   SREBP-1; SREBP-2; basic-helix-loop-helix-leucine zipper; plasma;
KW   transcription factor; low density lipoporotein; LDL; receptor;
KW   3-hydroxy-3-methylglutaryl coenzyme A synthase; cholesterol;
KW   hypercholesterolaemia; enhancer; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             118. .3543
FT                   /*tag= a
FT                   /product= SREBP-2
PN   WO9426922-A.
PD   24-NOV-1994.
PF   13-MAY-1994; U05300.
PR   13-MAY-1993; US-061697.
PR   01-OCT-1993; US-131365.
PA   (TEXA ) UNIV TEXAS SYSTEM.
PI   Briggs MR,  Brown MS,  Goldstein JL,  Wang X;
DR   WPI; 95-006813/01.
DR   P-PSDB; R66391.
PT   New sterol regulator element binding protein - used to develop
PT   prods. and screening assays for agents for reducing plasma
PT   cholesterol levels (Eng)
PS   Claim 34; Page 219-26; 305pp; English.
CC   The sequences given in Q79037-38 encode the sterol regulatory element
CC   binding proteins (SREBP), SREBP-1a and SREBP-2. SREBP's are regulatory
CC   proteins which are involved in the regulation of genes involved in
CC   cholesterol metabolism that are under the control of an associated
CC   sterol regulatory element-1 (SRE-1) enhancer sequence. SREBP proteins
CC   fall into two families, SREBP-1 and SREBP-2. Both proteins are members
CC   of a family of basic-helix-loop-helix-leucine zipper (bHLH-Zip)
CC   transcription factors. Each have the ability to bind to SRE sequences
CC   and modulate SRE-mediated transcription.  SRE-1 is a conditional enhancer
CC   found in the promoters for the low density lipoporotein (LDL) receptor
CC   and 3-hydroxy-3-methylglutaryl coenzyme A synthase genes.  It increases
CC   transcription in the absence of sterols and is inactivated when sterols
CC   accumulate.  Human SREBP-2 contins 1141 amino acids and has 47% identity
CC   with STREBP-1a.  SREBP-1a was the first recognised member of this family
CC   and has 1147 amino acids.  The resemblence between SREBP-1a and SREBP-2
CC   includes an acidic N-terminus, a highly conserved bHLH-Zip motif (71%
CC   identical), and an unusually long extension of 740 amino acids on the
CC   C-terminal side of the bHLH-Zip region.  SREBP-2 posseses one feature
CC   lacking in SREBP-1a, a Glu rich region (27% Glu over 121 residues).
CC   SREBP promotes SRE-1-mediated gene transcription, eg. LDL receptor
CC   production in the presence of sterols.  SREBP identified in screening
CC   assays, may be used to reduce plasma cholesterol levels and in
CC   controlling hypercholesterolaemia and its associated diseases.
SQ   Sequence   5197 BP;    1066 A;    1570 C;    1434 G;    1127 T;


  Query Match             3.0%;  Score 114.4;  DB 1;  Length 5197;
  Best Local Similarity   52.3%;  Pred. No. 1.7e-21;
  Matches  315;  Conservative    0;  Mismatches  266;  Indels   21;  Gaps    2;

Qy    2909 cctctgcgatgagtcgagtgtggagctccaagagtgcttgactgtcaaccggatcaccga 2968
           ||  |||||   | | ||||   | ||        || | |  ||||   ||  ||||
Db    2946 CCATTGCGAGAGGGCCAGTGGCCACCTATGGAGCAGCCTCAACGTCAGTGGGGGCACCTC 3005

Qy    2969 cgccaagggtataaagctgcttttccagttgcttacctgcgattggctgctcgaaactag 3028
           | |   |   | || |   |  ||||| |||| ||||| ||   ||| ||        |
Db    3006 TGACCCTGCCCTCAACCACGTGGTCCAGCTGCTCACCTGTGACCTGCTACTGTCGCTACG 3065

Qy    3029 gactgctctgtgggaactggaacacatgaatatggaggacgatgg---cttctaccaagt 3085
           ||| || || ||| ||   |   ||     |  ||| |  ||     ||||||
Db    3066 GACAGCGCTCTGGCAAAAACAGGCCAGTGCCAGCCAGGCTGTGGGGGAGACCTACCACGC 3125

Qy    3086 gccaggtgaagtgctcgagaagttccagaccgatttgaactcgttgcgcaacattgtgga 3145
           | |||| |  |  || | |   |||||    ||  ||  |    ||||||   | |   |
Db    3126 GTCAGGCGCTGAACTGGCGGGCTTCCAACGGGACCTGGGCAGCCTGCGCAGGCTGGCACA 3185

Qy    3146 gaatataccgaacgcccaatcgcgcatatatttgtacgaggcagtttgtcgcctgatggc 3205
           |   | |      ||  |         | |   || | || ||       ||||||||||
Db    3186 CAGCTTCCGCCCAGCATACCGCAAGGTGTTCCTGCATGAAGCCACCGTGCGCCTGATGGC 3245

Qy    3206 tggagcctcaccgtgtccaacgcaacagctcttggacaggagtctgcgatcacgcaacgc 3265
           |||| |   ||    |  || || |||||  ||||    || |||||    |||| | |
Db    3246 AGGAGGCAGCCCCACCCGCACCCACCAGCTGCTGGAACACAGCCTGCGGCGGCGCACCAC 3305

Qy    3266 ccactcgtccatcttctgcggcagcaaggatcggcggcagcagaacttcgtgggcggaga 3325
           ||                  ||||   | |   |  ||  | |   |  |  |||  |
Db    3306 GCA-----------------GAGCACCAAGCACGGAGAGGTGGATGCCTGGCCCGGCCA 3347

Qy    3326 gcgggaacgggcttcggccatgtacgtggcctgcaagtatctcccgcctgcgctgctcag 3385
           ||| || |||||  | |||||     |||||||     | || || | |   | | |||
Db    3348 GCGAGAGCGGGCCACCGCCATCCTGCTGGCCTGCCGCCACCTGCCCCTCTCCTTCCTCTC 3407

Qy    3386 ctccccgggtgaacgtgctggcatgttagccgaggcggccaagaccctggagaaggtggg 3445
           |||||||||  | || || |   || | ||||| || |||   ||||||||||||||||
Db    3408 CTCCCCGGGCCAGCGGGCAGTGCTGCTGGCCGAAGCTGCCCGCACCCTGGAGAAGGTGGG 3467

Qy    3446 cgacaagcgaaagctcaaggagtgctaccagctgatgaagtcgctgggcaacggcattgg 3505
           ||||  |||      ||| || ||| | ||| ||||     ||||||    |||| ||
Db    3468 CGACCGGCGCTCCTGCAACGACTGCCAGCAGATGATTGTTAAGCTGGGTGGTGGCACTGC 3527

Qy    3506 ca 3507
           ||
Db    3528 CA 3529


RESULT   3
Q84832
ID   Q84832 standard; DNA; 234 BP.
AC   Q84832;
DT   25-SEP-1995  (first entry)
DE   Spinocerebellar ataxia type 1 CAG repeat region patient #2.
KW   Spinocerebellar ataxia type 1; SCA 1; presymptomatic diagnosis;
KW   CAG repeat region; patient #2; ss.
OS   Homo sapiens.
PN   WO9501437-A.
PD   12-JAN-1995.
PF   29-JUN-1994; U07336.
PR   29-JUN-1993; US-084365.
PR   28-JUN-1994; US-267803.
```

```
PA   (MINU ) UNIV MINNESOTA.
PI   Chung M,  Orr HT,  Zoghbi HY;
DR   WPI; 95-061001/08.
PT   New autosomal dominant spinocerebellar ataxia type 1 nucleic acid
PT   - used to develop prods. for detection or presymptomatic
PT   diagnosis of a SCA1 disorder
PS   Disclosure; Fig 2; 111pp; English.
CC   Q84831-Q84835 show the CAG repeat regions of five individuals
CC   affected with spinocerebellar ataxia type 1 (SCA 1). It is within
CC   CAG repeat region (Q84804) that the mutations responsible for
CC   SCA 1 occur. The full nucleic acid (Q84793) and its protein product
CC   (R71111) can be used to develop products, for the presymptomatic
CC   detection of a SCA 1 disorder.
SQ   Sequence   234 BP;    73 A;    80 C;    77 G;    4 T;


  Query Match             1.3%;  Score 48.4;  DB 1;  Length 234;
  Best Local Similarity   51.9%;  Pred. No. 0.00062;
  Matches  109;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy     314 ggacctggcgccgacgcagatgtacaacatgctgctggacgagcctcgaacgcataccca 373
           | | | |  || |  |||| | |  |   || || | |  |||  |    |||    ||
Db      24 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 83

Qy     374 gcagacgcagtccgtggatcagcagccgcaatccgtcgagcaacagccgcacgtgaaaag 433
           ||||  ||||     | | ||||||| |||        |||| |||| |||   | |
Db      84 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 143

Qy     434 cgagcactcttcgccagtgcacatcaaggaggaactgcatcagcagcaacaacagtcgcc 493
             ||||      ||    |||     || || | | ||| |||||||| || |||  ||
Db     144 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 203

Qy     494 gcttctcgtctacaaaccagatcccctcat 523
           ||  | | ||  ||   |       |||||
Db     204 GCAGCACCTCAGCAGGGCTCCGGGGCTCAT 233


RESULT   4
Q98470/c
ID   Q98470 standard; cDNA; 2744 BP.
AC   Q98470;
DT   01-MAY-1996  (first entry)
DE   MiSP1-containing plasmid pMISS1.
KW   Spider silk; repeat unit; consensus; minor ampullate silk protein;
KW   spidroin; MiSP; orb web spider; dragline; ds.
OS   Nephila clavipes.
FH   Key             Location/Qualifiers
FT   cds             183. .2678
FT                   /*tag=  a
FT   misc_difference 996. .1037
FT                   /*tag= b
FT                   /note= "represented as indeterminate, a highly
FT                   compressed GC rich region which could
FT                   not be sequenced"
PN   WO9525165-A1.
PD   21-SEP-1995.
PF   14-MAR-1995; U03139.
PR   14-MAR-1994; US-209747.
PA   (UYWY-) UNIV WYOMING.
PI   Colgin M,  Lewis RV;
DR   WPI; 95-336970/43.
DR   P-PSDB; R80168.
PT   Polypeptide(s) comprising repeated unit amino acid sequences, also
PT   cDNAs - derived from minor ampullate spider silk proteins and used
PT   to form spider silk fibres
PS   Claim 12; Fig 1; 86pp; English.
CC   cDNA clone, pMISS1, encoding the orb web spider minor ampullate silk
CC   protein MiSP1 has been identified and sequenced. Repeat unit peptides
CC   of MiSP1 may form part of a larger polypeptide with an amino terminus
CC   (see R80184-85) and a carboxy terminus (see R80186-89). Repeat unit
CC   peptides make up spider silk proteins (spidroins) which in turn
CC   aggregate to form the silk fibres. Spider silk fibres have high
CC   tensile strength and significant elasticity. An isolated cDNA clone of
CC   a silk protein encoding sequence is of use to produce the protein at
CC   high yields using recombinant DNA technology.
SQ   Sequence   2744 BP;    486 A;    549 C;    1059 G;    608 T;


  Query Match             1.3%;  Score 47.2;  DB 1;  Length 2744;
  Best Local Similarity   50.0%;  Pred. No. 0.0059;
  Matches  118;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy     268 acgccgacctgcactcggacatgctggacatcatcctcaacgacatggacctggcgccga 327
           ||| | |||| |||   ||  || | | ||   |||    |  | | ||| ||| |
Db    1928 ACGACCACCTCTACTGTAACCCGCAGCGCCTCCAGCTCCTGCTCCAGCAGCTGCCGCAGC 1869

Qy     328 cgcagatgtacaacatgctgctggacgagcctcgaacgcatacccagcagacgcagtccg 387
           |||  | |    |  ||  |   | | |||       | || ||  |||   | |  |
Db    1868 ACCAGCTCTTGCTCCGGCACCGTATCCACCCTGACCACCGTAGCCTCCAGCTCCTGCTCC 1809

Qy     388 tggatcagcagccgcaatccgtcgagcaacagccgcacgtgaaaagcgagcactcttcgc 447
           || | | ||||| ||      || |||| | || |||  |  |   | |||||      |
Db    1808 TGCACCTGCAGCGGCTCCAGCTCCAGCACCTGCAGCAGCTCCAGCTCCAGCACCAGCACC 1749

Qy     448 cagtgcacatcaaggaggaactgcatcagcagcaacaacagtcgccgcttctcgtc 503
               | | || |    |||| | ||| || || |||  |||||  |   ||
Db    1748 TCTACCGTAACCAGCACCTCCTGCTCCTGCACCAGCAGCAGCAGCCGCACCAGCTC 1693


RESULT   5
V17226
ID   V17226 standard; DNA; 203 BP.
AC   V17226;
DT   29-JUN-1998  (first entry)
DE   SCA2 gene CAG repeat unit fragment.
KW   SCA2 gene; spinocerebellar ataxis type II; CAG repeat; PCR primer; ss.
OS   Synthetic.
PN   WO9803679-A1.
PD   29-JAN-1998.
PF   18-JUL-1996; J01999.
PR   18-JUL-1996; WO-J01999.
PA   (SRLS-) SRL INC.
PI   Sanpei K, Tsuji S;
DR   WPI; 98-120796/11.
PT   Diagnosing spinocerebellar ataxis type II - by PCR and determining
PT   number of CAG repeat units
PS   Example 1; Page 12; 23pp; Japanese.
CC   This sequence represents a fragment of the SCA2 gene. It can be used in
CC   the method of the invention for diagnosing spinocerebellar ataxis type
CC   II, by performing PCR on the test DNA using two primers hybridising to
CC   parts of the SCA2 gene sequence, and determining the number of CAG
CC   repeats in the amplified products. The method provides an easy means for
CC   the diagnosis of spinocerebellar ataxis type II.
SQ   Sequence   203 BP;    68 A;    70 C;    62 G;    3 T;


  Query Match             1.2%;  Score 47;  DB 1;  Length 203;
  Best Local Similarity   52.9%;  Pred. No. 0.0014;
  Matches  101;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy     316 acctggcgccgacgcagatgtacaacatgctgctggacgagcctcgaacgcatacccagc 375
           ||| |   | |   |||  | |  |   || || | |  |||  |    |||    ||||
Db       5 ACCAGCAACAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 64

Qy     376 agacgcagtccgtggatcagcagccgcaatccgtcgagcaacagccgcacgtgaaaagcg 435
           ||  ||||     | | ||||||| |||        ||||,|||| |||   | |
Db      65 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 124

Qy     436 agcactcttcgccagtgcacatcaaggaggaactgcatcagcagcaacaacagtcgccgc 495
           ||||      ||    |||     || || | | ||| ||||||| || |||  || ||
Db     125 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 184

Qy     496 ttctcgtctac 506
           || ||   ||
Db     185 ATCACGGAAAC 195
```

```
RESULT   6
V30271
ID   V30271 standard; DNA; 203 BP.
AC   V30271;
DT   02-OCT-1998  (first entry)
DE   Glutamine rich region encoding sequence found in SCA2 patients.
KW   Spinocerebellar ataxia type 2; SCA2; gene therapy; antisense therapy;
KW   CAG repeat; neurodegenerative disease; ss.
OS   Homo sapiens.
PN   WO9818920-A1.
PD   07-MAY-1998.
PF   30-OCT-1997; J03946.
PR   30-OCT-1996; JP-304059.
PA   (SRLS-) SRL INC.
PI   Sanpei K, Tsuji S;
DR   WPI; 98-272215/24.
PT   Nucleic acid fragments associated with spinocerebellar ataxia type 2
PT   - contain increased number of CAG repeat region compared to normal
PT   gene
PS   Example 1; Page 22; 38pp; Japanese.
CC   This represents a sequence encoding glutamine rich repeat region which
CC   can be found in spinocerebellar ataxia type 2 (SCA2) patients. The
CC   specification provides a gene sequence causative of the neurodegenerative
CC   disease SCA2, having a tri-nucleotide (CAG) repeat region which in the
CC   expression product produces a polyglutamine sequence from Gln-166 to
CC   Gln-188. In the normal gene there are 15-25 CAG repeats but in SCA2
CC   patients this number is increased to 35-100. Peptides encoded by nucleic
CC   acid fragments (DNA or RNA) containing sequences from the SCA2 associated
CC   gene, antibodies recognising the peptides and antisense nucleic acids
CC   hybridising with the nucleic acid fragments can be used for the
CC   investigation and diagnosis of SCA2. They can also be used for the
CC   treatment of SCA2 by antisense therapy or gene therapy.
SQ   Sequence   203 BP;    68 A;    70 C;    62 G;    3 T;


  Query Match             1.2%;  Score 47;  DB 1;  Length 203;
  Best Local Similarity   52.9%;  Pred. No. 0.0014;
  Matches  101;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy    316 acctggcgccgacgcagatgtacaacatgctgctggacgagcctcgaacgcatacccagc 375
          |||  |   | |   |||  | |  |   || || | |  |||  |    |||    ||||
Db      5 ACCAGCAACAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 64

Qy    376 agacgcagtccgtggatcagcagccgcaatccgtcgagcaacagccgcacgtgaaaagcg 435
          ||  ||||     | | ||||||| |||        |||| |||| |||   | |
Db     65 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 124

Qy    436 agcactcttcgccagtgcacatcaaggaggaactgcatcagcagcaacaacagtcgccgc 495
          ||||      ||    |||     || || | | ||| |||||||| || |||  || ||
Db    125 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 184

Qy    496 ttctcgtctac 506
           || ||   ||
Db    185 ATCACGGAAAC 195


RESULT   7
Q84831
ID   Q84831 standard; DNA; 195 BP.
AC   Q84831;
DT   25-SEP-1995  (first entry)
DE   Spinocerebellar ataxia type 1 CAG repeat region patient #1.
KW   Spinocerebellar ataxia type 1; SCA 1; presymptomatic diagnosis;
KW   CAG repeat region; patient #1; ss.
OS   Homo sapiens.
PN   WO9501437-A.
PD   12-JAN-1995.
PF   29-JUN-1994; U07336.
PR   29-JUN-1993; US-084365.
PR   28-JUN-1994; US-267803.
PA   (MINU ) UNIV MINNESOTA.
PI   Chung M,  Orr HT,  Zoghbi HY;
DR   WPI; 95-061001/08.
PT   New autosomal dominant spinocerebellar ataxia type 1 nucleic acid
PT   - used to develop prods. for detection or presymptomatic
PT   diagnosis of a SCA1 disorder
PS   Disclosure; Fig 2; 111pp; English.
CC   Q84831-Q84835 show the CAG repeat regions of five individuals
CC   affected with spinocerebellar ataxia type 1 (SCA 1). It is within
CC   CAG repeat region (Q84804) that the mutations responsible for
CC   SCA 1 occur. The full nucleic acid (Q84793) and its protein product
CC   (R71111) can be used to develop products, for the presymptomatic
CC   detection of a SCA 1 disorder.
SQ   Sequence   195 BP;    60 A;    67 C;    64 G;    4 T;


  Query Match             1.2%;  Score 45.8;  DB 1;  Length 195;
  Best Local Similarity   52.3%;  Pred. No. 0.003;
  Matches  101;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy    329 gcagatgtacaacatgctgctggacgagcctcgaacgcatacccagcagacgcagtccgt 388
          ||||  | |  |   || || | |  |||  |    |||    ||||||  ||||
Db      3 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 62

Qy    389 ggatcagcagccgcaatccgtcgagcaacagccgcacgtgaaaagcgagcactcttcgcc 448
          | | ||||||| |||        |||| |||| |||   | |     ||||      ||
Db     63 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 122

Qy    449 agtgcacatcaaggaggaactgcatcagcagcaacaacagtcgccgcttctcgtctacaa 508
             |||     || || | | ||| |||||||| || |||  || ||  |||  |
Db    123 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCACCTCAGCAGGGC 182

Qy    509 accagatcccctc 521
           || |  | | ||
Db    183 TCCGGGGCTCATC 195


RESULT   8
V73805/c
ID   V73805 standard; DNA; 32207 BP.
AC   V73805;
DT   25-FEB-1999  (first entry)
DE   KSHV LUR DNA (nucleotides 105,301-137,507).
KW   Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; DHFR; Bcl-2;
KW   dihydrofolate reductase; LUR; long unique region; vaccine; prophylaxis;
KW   diagnosis; treatment; HHV8; capsid protein IV; tegument protein IV;
KW   glycoprotein; kaposin; cyclin D; immediate early protein; IEP; OX-2;
KW   v-adh; G-protein coupled receptor; FGARAT; ds.
OS   Kaposi's sarcoma-associated herpesvirus.
PN   US5849564-A.
PD   15-DEC-1998.
PF   29-NOV-1996; 770379.
PR   29-NOV-1996; US-770379.
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
PI   Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;
DR   WPI; 99-069741/06.
PT   Kaposi's sarcoma-associated herpes virus nucleic acid - encodes
PT   di:hydro:folate reductase and is useful for treatment, prophylaxis
PT   or diagnosis of Kaposi's sarcoma
PS   Disclosure; Column 155-182; 109pp; English.
CC   This sequence is a fragment of the Kaposi's sarcoma-associated
CC   herpesvirus (KSHV) LUR (long unique region). This fragment contains
CC   coding regions for ORF65 which encodes capsid protein IV, ORF66, ORF67
CC   which encodes tegument protein IV, ORF68 which encodes a glycoprotein,
CC   ORF69, K12 which encodes kaposin, K13, ORF72 which encodes cyclin D,
CC   ORF73 which encodes immediate early protein (IEP), K14 which encodes
CC   OX-2 (v-adh), ORF74 which encodes G-protein coupled receptor, ORF75
CC   which encodes tegument protein/FGARAT, K15. KSHV is a new human
CC   Herpesvirus (HHV8) believed to cause Kaposi's sarcoma (KS) which is the
CC   most common form of neoplasm occurring in persons with acquired immune
CC   deficiency syndrome (AIDS). The DHFR protein is useful for vaccination,
CC   prophylaxis, diagnosis and treatment of a subject with Kaposi's sarcoma
CC   and for detecting expression of a DNA virus associated with Kaposi's
CC   sarcoma in a cell.
SQ   Sequence   32207 BP;    7229 A;    9156 C;    8713 G;    7109 T;
```

```
  Query Match             1.2%;  Score 44.8;  DB 1;  Length 32207;
  Best Local Similarity   53.4%;  Pred. No. 0.12;
  Matches   94;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy     321 gcgccgacgcagatgtacaacatgctgctggacgagcctcgaacgcatacccagcagacg 380
           | || |  ||||  | |  |   || || ||| ||||  |    |||     |||||  |
Db   19866 GAGCAGGAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAG 19807

Qy     381 cagtccgtggatcagcagccgcaatccgtcgagcaacagccgcacgtgaaaagcgagcac 440
           |||   |  || |||||||  |||     |   |||| |||| | | | | |      ||||
Db   19806 CAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAG 19747

Qy     441 tcttcgccagtgcacatcaaggaggaactgcatcagcagcaacaacagtcgccgct 496
             |   |  | |||     |||||  |   ||| ||| || | ||  ||  |  | |
Db   19746 GATGAACAGGAGCAGCAGGAGGAGCAGGAGCAGCAGGAGGAGCAGGAGCAGGAGTT 19691


RESULT   9
V19941/c
ID   V19941 standard; DNA; 137507 BP.
AC   V19941;
DT   03-AUG-1998  (first entry)
DE   KSHV long unique coding region and terminal repeat.
KW   KSHV; HHV8; human herpes virus 8; macrophage inflammatory protein II;
KW   interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis;
KW   complement-binding protein; glycoprotein; capsid protein IV; infection;
KW   immediate early protein; Kaposi's sarcoma; protective vaccine; lymphoma;
KW   lymphoproliferative disease; leukaemia; splenomegaly; mycosis fungoides;
KW   HIV immune status; anti-inflammatory agent; therapy; ds.
OS   Kaposi's sarcoma-associated herpes virus.
FH   Key             Location/Qualifiers
FT   CDS             1142. .2794
FT                   /*tag= a
FT                   /product= complement-binding protein
FT   CDS             8699. .11236
FT                   /*tag= b
FT                   /product= glycoprotein B
FT   CDS             complement (17261. .17875)
FT                   /*tag= c
FT                   /product= interleukin 6
FT   CDS             complement (21548. .21832)
FT                   /*tag= d
FT                   /product= macrophage inflammatory protein II
FT   CDS             complement (27137. .27424)
FT                   /*tag= e
FT                   /product= interferon regulatory factor 1
FT   CDS             28661. .29741
FT                   /*tag= f
FT                   /product= protein T1.1
FT   CDS             complement (58976. .60175)
FT                   /*tag= g
FT                   /product= glycoprotein M
FT   CDS             complement (69412. .69915)
FT                   /*tag= h
FT                   /product= glycoprotein L
FT   CDS             complement (88410. .88910)
FT                   /*tag= i
FT                   /product= interferon regulatory factor 2
FT   CDS             89600. .90541
FT                   /*tag= j
FT                   /product= interferon regulatory factor 3
FT   CDS             90173. .90643
FT                   /*tag= k
FT                   /product= glycoprotein X
FT   CDS             complement (93636. .94127)
FT                   /*tag= l
FT                   /product= interferon regulatory factor 4
FT   CDS             complement (111931. .112443)
FT                   /*tag= m
FT                   /product= capsid protein IV
FT   CDS             complement (123808. .127296)
FT                   /*tag= n
FT                   /product= immediate early protein
PN   WO9804576-A1.
PD   05-FEB-1998.
PF   22-JUL-1997; U13346.
PR   29-NOV-1996; US-757669.
PR   25-JUL-1996; US-686243.
PR   25-JUL-1996; US-686349.
PR   25-JUL-1996; US-686350.
PR   25-JUL-1996; US-687253.
PR   25-JUL-1996; US-688814.
PR   05-SEP-1996; US-708678.
PR   10-OCT-1996; US-728323.
PR   13-NOV-1996; US-747887.
PR   13-NOV-1996; US-748640.
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
PI   Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;
DR   WPI; 98-130615/12.
PT   New nucleic acid encoding Kaposi's sarcoma associated herpes virus
PT   proteins - useful for, e.g. detecting levels of HHV8 in, and
PT   preparation of vaccines for treatment of, HIV patients
PS   Example 2; Page 135-203; 230pp; English.
CC   This sequence represents the long unique region and terminal repeat of
CC   the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known
CC   as human herpes virus 8 (HHV8). This sequence contains the DNAs of the
CC   invention which encode KSHV polypeptides selected from: (a) viral
CC   macrophage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);
CC   (c) viral IRF 1; (d) complement-binding protein; glycoproteins B, M or L;
CC   (d) capsid protein IV encoded by ORF65; and (e) immediate early protein
CC   encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded
CC   by it, and antibodies (Ab) specific for the proteins are useful for
CC   detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body
CC   fluids or tissue samples. HHV8 infections can be treated with antisense
CC   or triplex forming molecules or agents that bind specifically to the
CC   protein. Ab may be used for prophylaxis or treatment of HHV8 infection,
CC   while the protein can be used in protective vaccines. Ab may also be used
CC   to differentiate between lymphomas, and HHV8 may be implicated in many
CC   other lymphoproliferative diseases such as lymphomas, leukaemia,
CC   splenomegaly and mycosis fungoides. Cells and animals containing the
CC   nucleic acid are useful for drug screening. HHV8-derived peptides can be
CC   used as targets for antiviral drugs, e.g. dihydrofolate reductase gene
CC   can be inhibited with methotrexate. These can also be used to determine
CC   the immune status of a patient infected with HIV. HHV8 derived protein
CC   viral MIP III may be used as an anti-inflammatory agent for,
CC   e.g. treating rheumatoid arthritis. This sequence is stated as containing
CC   81 open reading frames.
SQ   Sequence   137507 BP;    32579 A;    37795 C;    35758 G;    31375 T;


  Query Match             1.2%;  Score 44.8;  DB 1;  Length 137507;
  Best Local Similarity   53.4%;  Pred. No. 0.29;
  Matches   94;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy     321 gcgccgacgcagatgtacaacatgctgctggacgagcctcgaacgcatacccagcagacg 380
           | || |  ||||  | |  |   || || ||| ||||  |    |||     |||||  |
Db 125166 GAGCAGGAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAG 125107

Qy     381 cagtccgtggatcagcagccgcaatccgtcgagcaacagccgcacgtgaaaagcgagcac 440
           |||   |  || |||||||  |||     |   |||| |||| | | | | |      ||||
Db 125106 CAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAG 125047

Qy     441 tcttcgccagtgcacatcaaggaggaactgcatcagcagcaacaacagtcgccgct 496
             |   |  | |||     |||||  |   ||| ||| || | ||  ||  |  | |
Db 125046 GATGAACAGGAGCAGCAGGAGGAGCAGGAGCAGCAGGAGGAGCAGGAGCAGGAGTT 124991


RESULT  10
Q23092
ID   Q23092 standard; DNA; 543 BP.
AC   Q23092;
DT   17-AUG-1992  (first entry)
DE   Antigen tc-7a gene.
KW   Oocysts; MAb 12-07; sporozoite; ss.
```

```
OS   Eimeria tenella.
PN   WO9204460-A.
PD   19-MAR-1992.
PF   05-SEP-1991; U06430.
PR   12-SEP-1990; US-581693.
PA   (GENE-) GENEX CORP.
PI   Jacobson JW, Strausberg RL, Wilson SD, Pope SH, Strausberg SL;
PI   Ruff MD, Augustine PC, Danforth HD;
DR   WPI; 92-114365/14.
DR   P-PSDB; R22392.
PT   Vaccine against avian coccidiosis - comprising recombinant
PT   Eimeria antigen ac-1b or ac-6b gene, or microorganisms expressing
PT   them
PS   Claim 5; Page 39 + Fig 5; 56pp; English.
CC   To identify antigens of E. tenella, expression libraries were
CC   prepd. in lambda vector, lambda gt11, using cDNA prepd. from polyA
CC   mRNA isolated from E. tenella oocysts.  The cDNA expression
CC   library was screened with monoclonal antibody (MAb) 12-07 which was
CC   raised against the sporozoite stage of E. tenella.  The library
CC   to be screened was plated on a host that allows lysis and plaque
CC   formation.  During induction of the antigens encoded by the phage,
CC   the plaques were identified by screening the filters with Mab 12-07.
CC   The cDNA inserts from the MAb 12-07 positive phage were cloned into
CC   bacteriophage M13 and subjected to sequence analysis.  Following
CC   sequence analysis, E. tenella antigen tc-7a was identified.
SQ   Sequence   543 BP;    229 A;    135 C;    161 G;    18 T;


  Query Match             1.1%;  Score 42.2;  DB 1;  Length 543;
  Best Local Similarity   49.8%;  Pred. No. 0.055;
  Matches  107;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy    274 acctgcactcggacatgctggacatcatcctcaacgacatggacctggcgccgacgcaga 333
          | | |||   | |   || | |  |    |   || |  || | |||  ||    || |
Db    153 AACAGCAGCTGCAACAGCAGCAGCTACAGCAACACCAGCTGCAGCTGCAGCACCAGCTGC 212

Qy    334 tgtacaacatgctgctggacgagcctcgaacgcatacccagcagacgcagtccgtggatc 393
          | |  ||  ||||| | |  |||  |     ||    | ||||  |||     || | |
Db    213 AGCAGCACCAGCTGCAGCAGCAGCAGCAGCAACAGGAGCTGCAGCAGCACCAGCTGCAGC 272

Qy    394 agcagccgcaatccgtcgagcaacagccgcacgtgaaaagcgagcactcttcgccagtgc 453
           |||||||     |   ||||  ||| |||   | |  |  |||||     |  |  |
Db    273 GACAGCCGCAGCAGGAGCAGCAGGAGCAGCAAAGGCAGCGGAAGCACGAACAGAAACCGA 332

Qy    454 acatcaaggaggaactgcatcagcagcaacaacag 488
          |     || |||| | ||| ||| ||||   | ||
Db    333 AGGAGCAGCAGGAGCAGCAGCAGAAGCAGAGAAAG 367


RESULT  11
V30274
ID   V30274 standard; DNA; 165 BP.
AC   V30274;
DT   02-OCT-1998  (first entry)
DE   Glutamine rich region encoding sequence found in SCA2 patients.
KW   Spinocerebellar ataxia type 2; SCA2; gene therapy; antisense therapy;
KW   CAG repeat; neurodegenerative disease; ss.
OS   Homo sapiens.
PN   WO9818920-A1.
PD   07-MAY-1998.
PF   30-OCT-1997; J03946.
PR   30-OCT-1996; JP-304059.
PA   (SRLS-) SRL INC.
PI   Sanpei K, Tsuji S;
DR   WPI; 98-272215/24.
PT   Nucleic acid fragments associated with spinocerebellar ataxia type 2
PT   - contain increased number of CAG repeat region compared to normal
PT   gene
PS   Examples; Page 23; 38pp; Japanese.
CC   This represents a sequence encoding glutamine rich repeat region which
CC   can be found in spinocerebellar ataxia type 2 (SCA2) patients. The
CC   specification provides a gene sequence causative of the neurodegenerative
CC   disease SCA2, having a tri-nucleotide (CAG) repeat region which in the
CC   expression product produces a polyglutamine sequence from Gln-166 to
CC   Gln-188. In the normal gene there are 15-25 CAG repeats but in SCA2
CC   patients this number is increased to 35-100. Peptides encoded by nucleic
CC   acid fragments (DNA or RNA) containing sequences from the SCA2 associated
CC   gene, antibodies recognising the peptides and antisense nucleic acids
CC   hybridising with the nucleic acid fragments can be used for the
CC   investigation and diagnosis of SCA2. They can also be used for the
CC   treatment of SCA2 by antisense therapy or gene therapy.
SQ   Sequence   165 BP;    55 A;    55 C;    55 G;    0 U;


  Query Match             1.1%;  Score 41.8;  DB 1;  Length 165;
  Best Local Similarity   53.3%;  Pred. No. 0.035;
  Matches   88;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy    330 cagatgtacaacatgctgctggacgagcctcgaacgcatacccagcagacgcagtccgtg 389
          |||  | |  |   || || | |  |||  |    |||    ||||||  ||||    |
Db      1 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 60

Qy    390 gatcagcagccgcaatccgtcgagcaacagccgcacgtgaaaagcgagcactcttcgcca 449
          | |||||||  |||        ||||  ||||  |||   | |     ||||     ||
Db     61 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 120

Qy    450 gtgcacatcaaggaggaactgcatcagcagcaacaacagtcgccg 494
            |||     || || | | ||| ||||||||  || |||  || |
Db    121 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 165


RESULT  12
Q20685
ID   Q20685 standard; DNA; 14704 BP.
AC   Q20685;
DT   22-APR-1992  (first entry)
DE   PKS 741 insert contg. Sry.
KW   Y-chromosome; SRY; sex-determining region; genital ridge; ss.
OS   Mus musculus.
PN   WO9200375-A.
PD   09-JAN-1992.
PF   28-JUN-1991; G01057.
PR   28-JUN-1990; GB-014446.
PR   13-JUL-1990; GB-015488.
PR   09-MAY-1991; GB-010085.
PA   (IMCR ) IMP CANCER RES TECH.
PI   Goodfellow PN, Lovell-Badge R;
DR   WPI; 92-041557/05.
PT   Nucleic acid pY53.3, its fragments and oligo-nucleotide(s) - is
PT   used for determining or controlling sex of e.g. embryos by
PT   detecting Y chromosomes
PS   Claim 7; Page 122 and Fig 19(1-3); 183pp; English.
CC   The mouse gene equivalent to the human Y-located gene defined by
CC   pY53.3 (Q20684), is referred to as Sex-determining Region of the
CC   Y-chromosome (Sry) (formerly (mating type box y) Mty).
CC   The mouse Sry is expressed in male genital ridge, at the time that
CC   sex determination occurs, and adult testis.
CC   See also Q20680-81, Q20684-87, Q22530-31, Q22535-36 and Q22560-67.
SQ   Sequence   14704 BP;    4259 A;    3237 C;    3099 G;    4103 T;


  Query Match             1.1%;  Score 41.2;  DB 1;  Length 14704;
  Best Local Similarity   49.1%;  Pred. No. 0.76;
  Matches  109;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy    274 acctgcactcggacatgctggacatcatcctcaacgacatggacctggcgccgacgcaga 333
          |||  |||  | |    | | |  ||     | || ||  | | | |  || |  ||||
Db   8802 ACCACCACCAGCAGCAACAGCAGTTCTATGACCACCACCAGCAGCAGCAGCAGCAGC 8861

Qy    334 tgtacaacatgctgctggacgagcctcgaacgcatacccagcagacgcagtccgtggatc 393
           | |  |   || | |  | || |  |    ||| |  ||||||   ||   |    | |
Db   8862 AGCAGCAGCAGCAGTTCCATGACCACCACCAGCAGAAGCAGCAGTTTCATGACCACCACC 8921

Qy    394 agcagccgcaatccgtcgagcaacagccgcacgtgaaaagcgagcactcttcgccagtgc 453
          ||||||  ||     || |  | || |  |||    |    |||||    |  |  | |
Db   8922 AGCAGCAACAGCAGTTCCATGACCACCACCACCACCACCAGGAGCAGCAGTTCCATGACC 8981
```

```
Qy    454 acatcaaggaggaactgcatcagcagcaacaacagtcgccgc 495
          ||  | || || ||| |||    ||  | || |||  || ||
Db   8982 ACCACCAGCAGCAACAGCAGTTCCATGACCACCAGCAGCAGC 9023


RESULT  13
Q84834
ID   Q84834 standard; DNA; 171 BP.
AC   Q84834;
DT   25-SEP-1995  (first entry)
DE   Spinocerebellar ataxia type 1 CAG repeat region patient #4.
KW   Spinocerebellar ataxia type 1; SCA 1; presymptomatic diagnosis;
KW   CAG repeat region; patient #4; ss.
OS   Homo sapiens.
PN   WO9501437-A.
PD   12-JAN-1995.
PF   29-JUN-1994; U07336.
PR   29-JUN-1993; US-084365.
PR   28-JUN-1994; US-267803.
PA   (MINU ) UNIV MINNESOTA.
PI   Chung M,  Orr HT,  Zoghbi HY;
DR   WPI; 95-061001/08.
PT   New autosomal dominant spinocerebellar ataxia type 1 nucleic acid
PT   - used to develop prods. for detection or presymptomatic
PT   diagnosis of a SCA1 disorder
PS   Disclosure; Fig 2; 111pp; English.
CC   Q84831-Q84835 show the CAG repeat regions of five individuals
CC   affected with spinocerebellar ataxia type 1 (SCA 1). It is within
CC   CAG repeat region (Q84804) that the mutations responsible for
CC   SCA 1 occur. The full nucleic acid (Q84793) and its protein product
CC   (R71111) can be used to develop products, for the presymptomatic
CC   detection of a SCA 1 disorder.
SQ   Sequence   171 BP;    52 A;    59 C;    56 G;    4 T;


  Query Match             1.1%;  Score 41;  DB 1;  Length 171;
  Best Local Similarity   52.7%;  Pred. No. 0.059;
  Matches   89;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy    355 agcctcgaacgcatacccagcagacgcagtccgtggatcagcagccgcaatccgtcgagc 414
          |||  |    |||    ||||||  ||||     | | |||||||  |||        |||
Db      2 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 61

Qy    415 aacagccgcacgtgaaaagcgagcactcttcgccagtgcacatcaaggaggaactgcatc 474
          | |||| |||   | |     ||||       ||    |||     || || | | ||| |
Db     62 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 121

Qy    475 agcagcaacaacagtcgccgcttctcgtctacaaaccagatcccctcat 523
          ||||||| || |||  || ||  | | ||  ||   |       |||||
Db    122 AGCAGCAGCAGCAGCAGCAGCACCTCAGCAGGGCTCCGGGGCTCAT 170


RESULT  14
X23749
ID   X23749 standard; DNA; 1498 BP.
AC   X23749;
DT   25-JUN-1999  (first entry)
DE   Rice anthranilate synthase first isozyme alpha-subunit DNA fragment.
KW   Anthranilate synthase; alpha-subunit; ASA; rice; isozyme; plant; seed;
KW   maize; wheat; tryptophan content; nutritional value; ss.
OS   Oryza sativa.
FH   Key             Location/Qualifiers
FT   CDS             1102. .1498
FT                   /*tag=  a
FT                   /product= "ASA synthase"
FT                   /note= "Partial coding sequence. Incomplete exon 2"
FT   exon            1102. .1233
FT                   /*tag=  b
FT                   /number= 1
FT   intron          1234. .1318
FT                   /*tag=  c
FT                   /number= 1
FT   exon            1318. .1498
FT                   /*tag=  d
FT                   /number= 2
FT                   /note= "partial exon 2 sequence"
PN   WO9911800-A1.
PD   11-MAR-1999.
PF   31-AUG-1998; J03883.
PR   29-AUG-1997; JP-235049.
PA   (HOKK ) HOKKO CHEM IND CO LTD.
PA   (NORQ ) JAPAN MIN AGRIC FORESTRY & FISHERIES.
PI   Hasegawa H, Terakawa T, Tozawa Y, Wakasa K;
DR   WPI; 99-228982/19.
DR   P-PSDB; W93811.
PT   DNA encodes ^a-subunit of first isozyme of rice anthranilate
PT   synthase - used for improving tryptophan production and
PT   nutritional value of crops, e.g. rice, maize or wheat
PS   Example 5; Page 131-133; 152pp; Japanese.
CC   This invention describes a novel rice anthranilate synthase first
CC   isozyme alpha-subunit. The encoding DNA can be used to produce
CC   transformant plants and seeds, of e.g. rice, maize or wheat, with
CC   enhanced tryptophan content and nutritional value of the crops.
SQ   Sequence   1498 BP;    343 A;    362 C;    404 G;    389 T;


  Query Match             1.1%;  Score 41;  DB 1;  Length 1498;
  Best Local Similarity   54.2%;  Pred. No. 0.22;
  Matches   83;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy    552 cagccgacgggccttttgaaggccgcccaaccaacagccaccatacatcacatggacgcc 611
          | | || ||  ||| || ||   || ||       | ||   |   || |      |  ||
Db   1003 CGGGCGCCGCACCTATTTAAATGCGGCCCCGATCCCGCATTCTCTCAACTGCACTAGTCC 1062

Qy    612 cagcggatgccgccgaacacggcggtgtatcccccatctctgggcagtagctttgtctac 671
          |  |  | | | | |  ||     || |||||||||| || ||| |||   | |  ||| |
Db   1063 CCACCAACGGCTCGGTCCAGTAGAGTTTATCCCCCACCTATGGCCAGCCTCGTGCTCTCC 1122

Qy    672 cagtccatgtccccgcccacgtcgccggtggag 704
          | |  |||  |||   ||||| ||||| ||| |
Db   1123 CTGCGCATCGCCCGTTCCACGCCGCCGCTGGGG 1155


RESULT  15
Q84835
ID   Q84835 standard; DNA; 154 BP.
AC   Q84835;
DT   25-SEP-1995  (first entry)
DE   Spinocerebellar ataxia type 1 CAG repeat region patient #5.
KW   Spinocerebellar ataxia type 1; SCA 1; presymptomatic diagnosis;
KW   CAG repeat region; patient #5; ss.
OS   Homo sapiens.
PN   WO9501437-A.
PD   12-JAN-1995.
PF   29-JUN-1994; U07336.
PR   29-JUN-1993; US-084365.
PR   28-JUN-1994; US-267803.
PA   (MINU ) UNIV MINNESOTA.
PI   Chung M,  Orr HT,  Zoghbi HY;
DR   WPI; 95-061001/08.
PT   New autosomal dominant spinocerebellar ataxia type 1 nucleic acid
PT   - used to develop prods. for detection or presymptomatic
PT   diagnosis of a SCA1 disorder
PS   Disclosure; Fig 2; 111pp; English.
CC   Q84831-Q84835 show the CAG repeat regions of five individuals
CC   affected with spinocerebellar ataxia type 1 (SCA 1). It is within
CC   CAG repeat region (Q84804) that the mutations responsible for
CC   SCA 1 occur. The full nucleic acid (Q84793) and its protein product
CC   (R71111) can be used to develop products, for the presymptomatic
CC   detection of a SCA 1 disorder.
SQ   Sequence   154 BP;    51 A;    50 C;    52 G;    1 T;


  Query Match             1.1%;  Score 40.6;  DB 1;  Length 154;
  Best Local Similarity   54.3%;  Pred. No. 0.072;
```

Matches 82; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

```
Qy     344 gctgctggacgagcctcgaacgcatacccagcagacgcagtccgtggatcagcagccgca 403
           || || | |  ||| |    |||    |||||  ||||     | | |||||| |||
Db       4 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 63

Qy     404 atccgtcgagcaacagccgcacgtgaaaagcgagcactcttcgccagtgcacatcaagga 463
             |||| |||| |||   | |     ||||      ||    |||     || |
Db      64 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 123

Qy     464 ggaactgcatcagcagcaacaacagtcgccg 494
           | | | ||| |||||||| || ||| || |
Db     124 GCAGCAGCAGCAGCAGCAGCAGCAGCAG 154
```

Search completed: September 4, 2000, 01:05:08
Job time: 18483 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2000, 00:33:20 ; Search time 183.89 Seconds
(without alignments)
2818.589 Million cell updates/sec

Title: US-09-332-522B-7
Perfect score: 3768
Sequence: 1 gtttattaagctgcaaatat.....aatattttaacagatttaaa 3768

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

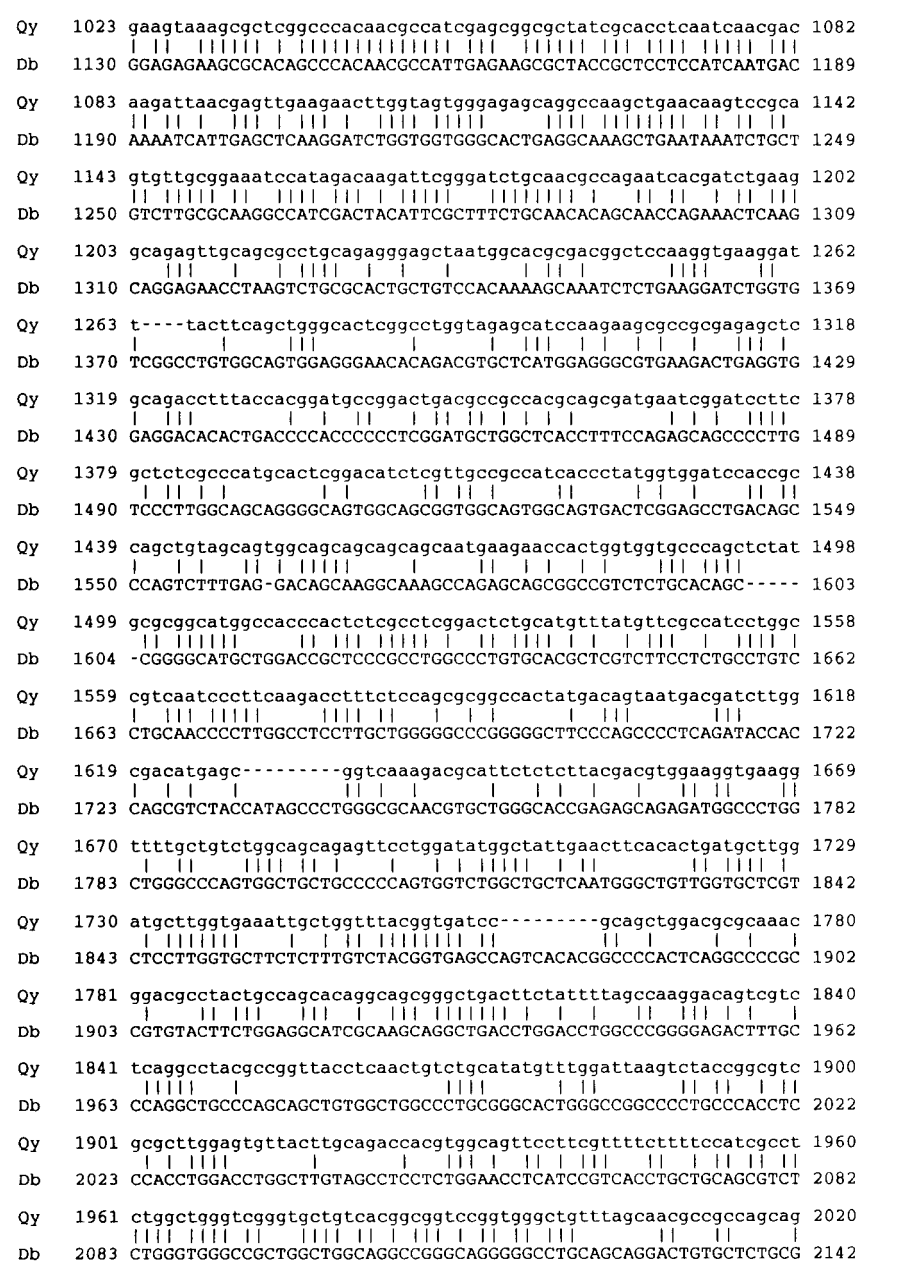
SUMMARIES

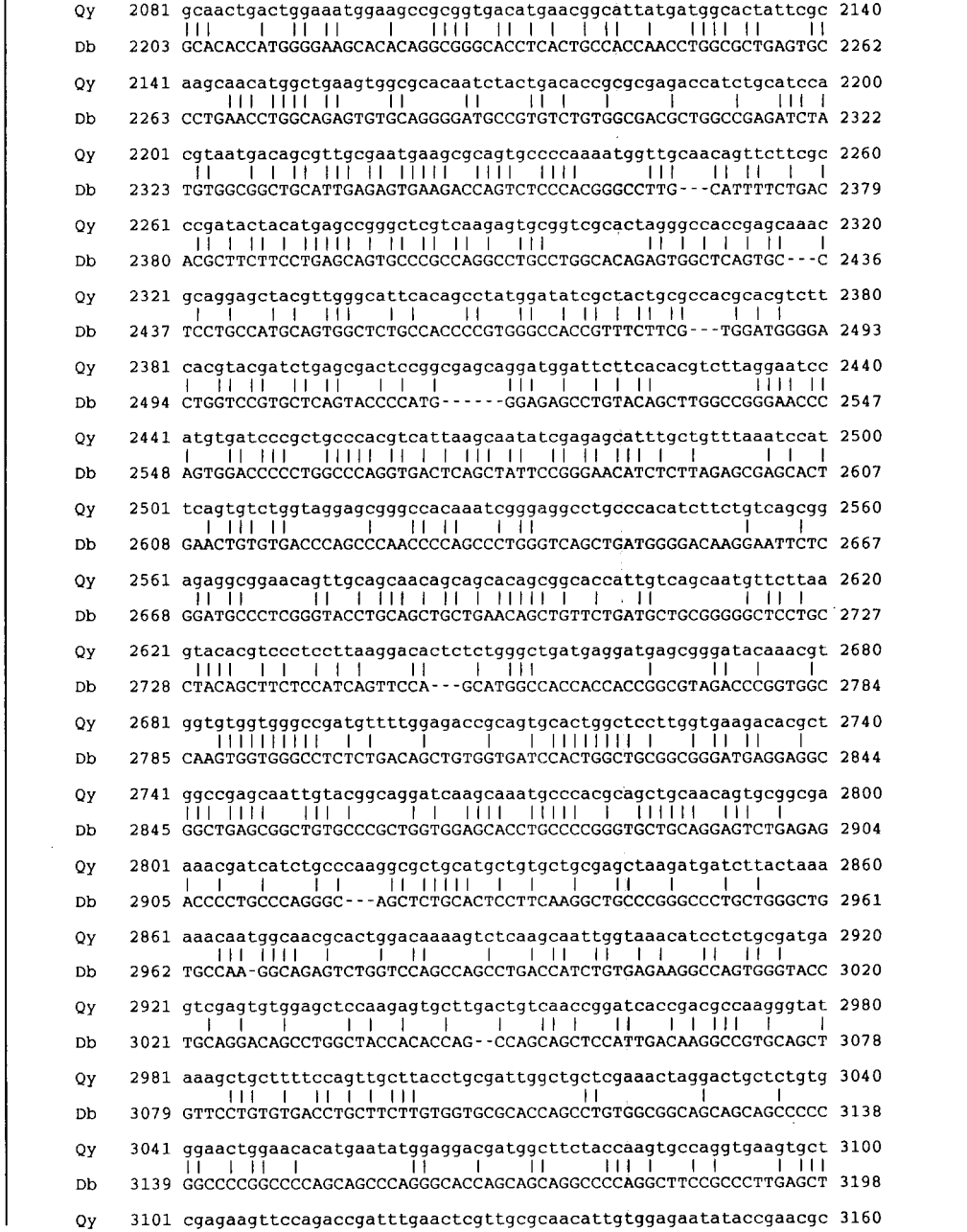
	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	165	4.4	4154	1	US-08-131-365B-37	Sequence 37, Appl
	2	165	4.4	4154	3	US-08-668-123-37	Sequence 37, Appl
	3	114.4	3.0	5197	1	US-08-131-365B-53	Sequence 53, Appl
	4	114.4	3.0	5197	3	US-08-668-123-53	Sequence 53, Appl
	5	48.4	1.3	234	2	US-08-469-802B-3	Sequence 3, Appli
	6	48.4	1.3	234	3	US-08-267-803B-3	Sequence 3, Appli
c	7	47.2	1.3	2793	2	US-08-209-747-1	Sequence 1, Appli
c	8	47.2	1.3	2793	2	US-08-458-298-1	Sequence 1, Appli
	9	45.8	1.2	195	2	US-08-469-802B-2	Sequence 2, Appli
	10	45.8	1.2	195	3	US-08-267-803B-2	Sequence 2, Appli
	11	44.8	1.2	3489	4	US-08-728-323A-1	Sequence 1, Appli
c	12	44.8	1.2	32207	3	US-08-770-379-20	Sequence 20, Appl
c	13	44	1.2	7218	1	US-08-232-463-14	Sequence 14, Appl
	14	42.2	1.1	543	7	5273901-6	Patent No. 5273901
	15	41.6	1.1	7218	1	US-08-232-463-14	Sequence 14, Appl
	16	41	1.1	171	2	US-08-469-802B-5	Sequence 5, Appli
	17	41	1.1	171	3	US-08-267-803B-5	Sequence 5, Appli
	18	40.6	1.1	154	2	US-08-469-802B-6	Sequence 6, Appli
	19	40.6	1.1	154	3	US-08-267-803B-6	Sequence 6, Appli
	20	40.6	1.1	168	2	US-08-469-802B-4	Sequence 4, Appli
	21	40.6	1.1	168	3	US-08-267-803B-4	Sequence 4, Appli
	22	40	1.1	533	7	5482709-5	Patent No. 5482709
c	23	38.8	1.0	786	2	US-08-403-379A-2	Sequence 2, Appli
c	24	38.8	1.0	786	4	US-08-929-414-2	Sequence 2, Appli
	25	38.8	1.0	793	4	US-09-018-595B-1	Sequence 1, Appli
	26	38.8	1.0	1187	2	US-08-440-856A-2	Sequence 2, Appli
c	27	38.2	1.0	3252	4	US-08-809-740A-1	Sequence 1, Appli
c	28	38.2	1.0	3252	4	US-08-809-740A-4	Sequence 4, Appli
c	29	38.2	1.0	50341	2	US-08-247-901C-1	Sequence 1, Appli
c	30	38.2	1.0	50341	4	US-09-075-904-1	Sequence 1, Appli
	31	38	1.0	3366	2	US-08-469-802B-1	Sequence 1, Appli
	32	38	1.0	3366	3	US-08-267-803B-1	Sequence 1, Appli
	33	38	1.0	10660	3	US-08-267-803B-8	Sequence 8, Appli
	34	37.2	1.0	506	2	US-08-469-802B-7	Sequence 7, Appli
	35	37.2	1.0	506	3	US-08-267-803B-7	Sequence 7, Appli
	36	37.2	1.0	1519	1	US-07-971-759-19	Sequence 19, Appl
	37	36.6	1.0	228	7	5273901-10	Patent No. 5273901
	38	36.6	1.0	228	7	5482709-9	Patent No. 5482709
	39	36.6	1.0	2287	1	US-08-222-619-1	Sequence 1, Appli
	40	36.6	1.0	2287	6	PCT-US95-04075-1	Sequence 1, Appli
	41	36.4	1.0	1995	2	US-08-425-069-3	Sequence 3, Appli
	42	36.4	1.0	1995	4	US-08-317-844B-3	Sequence 3, Appli
	43	36.2	1.0	802	4	US-09-018-595B-2	Sequence 2, Appli
c	44	36.2	1.0	4615	1	US-08-188-582-1	Sequence 1, Appli
c	45	36.2	1.0	4615	1	US-08-646-715-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-08-131-365B-37
; Sequence 37, Application US/08131365B
; Patent No. 5527690
;  GENERAL INFORMATION:
;    APPLICANT:  Brown, Michael S.
;    APPLICANT:  Briggs, Michael R.
;    APPLICANT:  Wang, Xiaodong
;    APPLICANT:  Goldstein, Joseph L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS RELATING
;    TITLE OF INVENTION:  TO STEROL REGULATORY ELEMENT BINDING
;    TITLE OF INVENTION:  PROTEINS
;    NUMBER OF SEQUENCES:  64
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  U.S.A.
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/131,365B
;      FILING DATE:  01-OCT-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  UTSD:372/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4154 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "DNA"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  167..3607
US-08-131-365B-37
```

Query Match 4.4%; Score 165; DB 1; Length 4154;
Best Local Similarity 45.7%; Pred. No. 8.7e-38;
Matches 1144; Conservative 0; Mismatches 1295; Indels 62; Gaps 14;





```
Db   3199 GCGTGGCTTCCAACGGGACCTGAGCAGCCTGAGGCGGCTGGCACAGAGCTTCCGGCCCGC 3258

Qy   3161 ccaatcgcgcatatatttgtacgaggcagtttgtcgcctgatggctggagcctcaccgtg 3220
Db   3259 CATGCGGAGGGTGTTCCTACATGAGGCCACGGCCCGGCTGATGGCGGGGGCCAGCCCCAC 3318

Qy   3221 tccaacgcaacagctcttggacaggagtctgcgatcacgcaacgcccactcgtccatctt 3280
Db   3319 ACGGACACACCAGCTCCTCGACCGCAGTCTGAGGCGGCGGGCAGGCCCCGGTGGCAAAGG 3378

Qy   3281 ctgcggcagcaaggatcggcggcagcagaacttcgtgggcggagagcgggaacgggcttc 3340
Db   3379 AGGCGCGGTGGCGGAGCTGGAGCCGC---------GGCCCACGCGGCGGGAGCACGCGGA 3429

Qy   3341 ggccatgtacgtggcctgcaagtatctcccgcctgcgctgctcagctccccgggtgaacg 3400
Db   3430 GGCCTTGCTGCTGGCCTCCTGCTACCTGCCCCCCGGCTTCCTGTCGGCGCCCGGGCAGCG 3489

Qy   3401 tgctggcatgttagccgaggcggccaagaccctggagaaggtgggcgacaagcgaaagct 3460
Db   3490 CGTGGGCATGCTGGCTGAGGCGGCGCGCACACTCGAGAAGCTTGGCGATCGCCGGCTGCT 3549

Qy   3461 caaggagtgctaccagctgatgaagtcgctgggcaacggca 3501
Db   3550 GCACGACTGTCAGCAGATGCTCATGCGCCTGGGCGGTGGGA 3590


RESULT   2
US-08-668-123-37
; Sequence 37, Application US/08668123
; Patent No. 5891631
;  GENERAL INFORMATION:
;    APPLICANT:  Brown, Michael S.
;    APPLICANT:  Briggs, Michael R.
;    APPLICANT:  Wang, Xiaodong
;    APPLICANT:  Goldstein, Joseph L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS RELATING
;    TITLE OF INVENTION:  TO STEROL REGULATORY ELEMENT BINDING
;    TITLE OF INVENTION:  PROTEINS
;    NUMBER OF SEQUENCES:  64
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  U.S.A.
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/668,123
;      FILING DATE:  14-JUN-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/131,365
;      FILING DATE:  01-OCT-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  UTSD:372/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4154 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "DNA"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  167..3607
US-08-668-123-37


  Query Match             4.4%;  Score 165;  DB 3;  Length 4154;
  Best Local Similarity   45.7%;  Pred. No. 8.7e-38;
  Matches 1144;  Conservative    0;  Mismatches 1295;  Indels   62;  Gaps   14;

Qy   1023 gaagtaaagcgctcggcccacaacgccatcgagcggcgctatcgcacctcaatcaacgac 1082
Db   1130 GGAGAGAAGCGCACAGCCCACAACGCCATTGAGAAGCGCTACCGCTCCTCCATCAATGAC 1189

Qy   1083 aagattaacgagttgaagaacttggtagtgggagagcaggccaagctgaacaagtccgca 1142
Db   1190 AAAATCATTGAGCTCAAGGATCTGGTGGTGGGCACTGAGGCAAAGCTGAATAAATCTGCT 1249

Qy   1143 gtgttgcggaaatccatagacaagattcgggatctgcaacgccagaatcacgatctgaag 1202
Db   1250 GTCTTGCGCAAGGCCATCGACTACATTCGCTTTCTGCAACACAGCAACCAGAAACTCAAG 1309

Qy   1203 gcagagttgcagcgcctgcagagggagctaatggcacgcgacggctccaaggtgaaggat 1262
Db   1310 CAGGAGAACCTAAGTCTGCGCACTGCTGTCCACAAAAGCAAATCTCTGAAGGATCTGGTG 1369

Qy   1263 t----tacttcagctgggcactcggcctggtagagcatccaagaagcgccgcgagagctc 1318
Db   1370 TCGGCCTGTGGCAGTGGAGGGAACACAGACGTGCTCATGGAGGGCGTGAAGACTGAGGTG 1429

Qy   1319 gcagacctttaccacggatgccggactgacgccgccacgcagcgatgaatcggatccttc 1378
Db   1430 GAGGACACACTGACCCCACCCCCCTCGGATGCTGGCTCACCTTTCCAGAGCAGCCCCTTG 1489

Qy   1379 gctctcgcccatgcactcggacatctcgttgccgccatcaccctatggtggatccaccgc 1438
Db   1490 TCCCTTGGCAGCAGGGGCAGTGGCAGCGGTGGCAGTGGCAGTGACTCGGAGCCTGACAGC 1549

Qy   1439 cagctgtagcagtggcagcagcagcaatgaagaaccactggtggtgcccagctctat 1498
Db   1550 CCAGTCTTTGAG-GACAGCAAGGCAAAGCCAGAGCAGCGGCCGTCTCTGCACAGC----- 1603

Qy   1499 gcgcggcatggccacccactctcgcctcggactctgcatgtttatgttcgccatcctggc 1558
Db   1604 -CGGGGCATGCTGGACCGCTCCCGCCTGGCCCTGTGCACGCTCGTCTTCCTCTGCCTGTC 1662

Qy   1559 cgtcaatcccttcaagacctttctccagcgcggccactatgacagtaatgacgatcttgg 1618
Db   1663 CTGCAACCCCTTGGCCTCCTTGCTGGGGGCCCGGGGGCTTCCCAGCCCCTCAGATACCAC 1722

Qy   1619 cgacatgagc---------ggtcaaagacgcattctctcttacgacgtggaaggtgaagg 1669
Db   1723 CAGCGTCTACCATAGCCCTGGGCGCAACGTGCTGGGCACCGAGAGCAGAGATGGCCCTGG 1782

Qy   1670 ttttgctgtctggcagcagagttcctggatatggctattgaacttcacactgatgcttgg 1729
Db   1783 CTGGGCCCAGTGGCTGCTGCCCCCAGTGGTCTGGCTGCTCAATGGGCTGTTGGTGCTCGT 1842

Qy   1730 atgcttggtgaaaattgctggtttacggtgatcc---------gcagctggacgcgcaaac 1780
Db   1843 CTCCTTGGTGCTTCTCTTTGTCTACGGTGAGCCAGTCACACGGCCCCACTCAGGCCCCGC 1902

Qy   1781 ggacgcctactgccagcacaggcagcgggctgacttctatttttagccaaggacagtcgtc 1840
Db   1903 CGTGTACTTCTGGAGGCATCGCAAGCAGGCTGACCTGGACCTGGCCCGGGGAGACTTTGC 1962

Qy   1841 tcaggcctacgccggttacctcaactgtctgcatatgtttggattaagtctaccggcgtc 1900
Db   1963 CCAGGCTGCCCAGCAGCTGTGGCTGGCCCTGCGGGCACTGGGCCGGCCCCTGCCCACCTC 2022

Qy   1901 gcgcttggagtgttacttgcagaccacgtggcagttccttcgttttctttccatcgcct 1960
```

```
Db   2023 CCACCTGGACCTGGCTTGTAGCCTCCTCTGGAACCTCATCCGTCACCTGCTGCAGCGTCT 2082

Qy   1961 ctggctgggtcgggtgctgtcacggcggtccggtgggctgtttagcaacgccgccagcag 2020

Db   2083 CTGGGTGGGCCGCTGGCTGGCAGGCCGGGCAGGGGGCCTGCAGCAGGACTGTGCTCTGCG 2142

Qy   2021 gaaacaggcgctggcatctgcacgcgaactggccctgctcttcaaccgactgaatcaatt 2080

Db   2143 AGTGGATGCTAGCGCCAGCGCCCGAGACGCAGCCCTGGTCTACCATAAGCTGCACCAGCT 2202

Qy   2081 gcaactgactggaaatggaagccgcggtgacatgaacggcattatgatggcactattcgc 2140

Db   2203 GCACACCATGGGGAAGCACACAGGCGGGCACCTCACTGCCACCAACCTGGCGCTGAGTGC 2262

Qy   2141 aagcaacatggctgaagtggcgcacaatctactgacaccgcgcgagaccatctgcatcca 2200

Db   2263 CCTGAACCTGGCAGAGTGTGCAGGGGATGCCGTGTCTGTGGCGACGCTGGCCGAGATCTA 2322

Qy   2201 cgtaatgacagcgttgcgaatgaagcgcagtgccccaaaatggttgcaacagttcttcgc 2260

Db   2323 TGTGGCGGCTGCATTGAGAGTGAAGACCAGTCTCCCACGGGCCTTG---CATTTTCTGAC 2379

Qy   2261 ccgatactacatgagccgggctcgtcaagagtgcggtcgcactagggccaccgagcaaac 2320

Db   2380 ACGCTTCTTCCTGAGCAGTGCCCGCCAGGCCTGCCTGGCACAGAGTGGCTCAGTGC---C 2436

Qy   2321 gcaggagctacgttgggcattcacagcctatggatatcgctactgcgccacgcacgtctt 2380

Db   2437 TCCTGCCATGCAGTGGCTCTGCCACCCCGTGGGCCACCGTTTCTTCG---TGGATGGGGA 2493

Qy   2381 cacgtacgatctgagcgactccggcgagcaggatggattcttcacacgtcttaggaatcc 2440

Db   2494 CTGGTCCGTGCTCAGTACCCCATG------GGAGAGCCTGTACAGCTTGGCCGGGAACCC 2547

Qy   2441 atgtgatcccgctgcccacgtcattaagcaatatcgagagcatttgctgtttaaatccat 2500

Db   2548 AGTGGACCCCCTGGCCCAGGTGACTCAGCTATTCCGGGAACATCTCTTAGAGCGAGCACT 2607

Qy   2501 tcagtgtctggtaggagcgggccacaaatcgggaggcctgcccacatcttctgtcagcgg 2560

Db   2608 GAACTGTGTGACCCAGCCCAACCCCAGCCCTGGGTCAGCTGATGGGGACAAGGAATTCTC 2667

Qy   2561 agaggcggaacagttgcagcaacagcagcacagcggcaccattgtcagcaatgttcttaa 2620

Db   2668 GGATGCCCTCGGGTACCTGCAGCTGCTGAACAGCTGTTCTGATGCTGCGGGGGCTCCTGC 2727

Qy   2621 gtacacgtccctccttaaggacactctctgggctgatgaggatgagcgggatacaaacgt 2680

Db   2728 CTACAGCTTCTCCATCAGTTCCA---GCATGGCCACCACCACCGGCGTAGACCCGGTGGC 2784

Qy   2681 ggtgtggtgggccgatgttttggagaccgcagtgcactggctccttggtgaagacacgct 2740

Db   2785 CAAGTGGTGGGCCTCTCTGACAGCTGTGGTGATCCACTGGCTGCGGCGGGATGAGGAGGC 2844

Qy   2741 ggccgagcaattgtacggcaggatcaagcaaatgcccacgcagctgcaacagtgcggcga 2800

Db   2845 GGCTGAGCGGCTGTGCCCGCTGGTGGAGCACCTGCCCCGGGTGCTGCAGGAGTCTGAGAG 2904

Qy   2801 aaacgatcatctgcccaaggcgctgcatgctgtgctgcgagctaagatgatcttactaaa 2860

Db   2905 ACCCCTGCCCAGGGC---AGCTCTGCACTCCTTCAAGGCTGCCCGGGCCCTGCTGGGCTG 2961

Qy   2861 aaacaatggcaacgcactggacaaaagtctcaagcaattggtaaacatcctctgcgatga 2920

Db   2962 TGCCAA-GGCAGAGTCTGGTCCAGCCAGCCTGACCATCTGTGAGAAGGCCAGTGGGTACC 3020

Qy   2921 gtcgagtgtggagctccaagagtgcttgactgtcaaccggatcaccgacgccaagggtat 2980

Db   3021 TGCAGGACAGCCTGGCTACCACACCAG--CCAGCAGCTCCATTGACAAGGCCGTGCAGCT 3078

Qy   2981 aaagctgcttttccagttgcttacctgcgattggctgctcgaaactaggactgctctgtg 3040

Db   3079 GTTCCTGTGTGACCTGCTTCTTGTGGTGCGCACCAGCCTGTGGCGGCAGCAGCAGCCCCC 3138

Qy   3041 ggaactggaacacatgaatatggaggacgatggcttctaccaagtgccaggtgaagtgct 3100

Db   3139 GGCCCCGGCCCCAGCAGCCCAGGGCACCAGCAGCAGGCCCCAGGCTTCCGCCCTTGAGCT 3198

Qy   3101 cgagaagttccagaccgatttgaactcgttgcgcaacattgtggagaatataccgaacgc 3160

Db   3199 GCGTGGCTTCCAACGGGACCTGAGCAGCCTGAGGCGGCTGGCACAGAGCTTCCGGCCCGC 3258

Qy   3161 ccaatcgcgcatatatttgtacgaggcagtttgtcgcctgatggctggagcctcaccgtg 3220

Db   3259 CATGCGGAGGGTGTTCCTACATGAGGCCACGGCCCGGCTGATGGCGGGGGCCAGCCCCAC 3318

Qy   3221 tccaacgcaacagctcttggacaggagtctgcgatcacgcaacgcccactcgtccatctt 3280

Db   3319 ACGGACACACCAGCTCCTCGACCGCAGTCTGAGGCGGCGGGCAGGCCCCGGTGGCAAAGG 3378

Qy   3281 ctgcggcagcaaggatcggcggcagcagaacttcgtgggcggagagcgggaacgggcttc 3340

Db   3379 AGGCGCGGTGGCGGAGCTGGAGCCGC---------GGCCCACGCGGCGGGAGCACGCGGA 3429

Qy   3341 ggccatgtacgtggcctgcaagtatctcccgcctgcgctgctcagctccccgggtgaacg 3400

Db   3430 GGCCTTGCTGCTGGCCTCCTGCTACCTGCCCCCCGGCTTCCTGTCGGCGCCCGGGCAGCG 3489

Qy   3401 tgctggcatgttagccgaggcggccaagaccctggagaaggtgggcgacaagcgaaagct 3460

Db   3490 CGTGGGCATGCTGGCTGAGGCGGCGCGCACACTCGAGAAGCTTGGCGATCGCCGGCTGCT 3549

Qy   3461 caaggagtgctaccagctgatgaagtcgctgggcaacggca 3501

Db   3550 GCACGACTGTCAGCAGATGCTCATGCGCCTGGGCGGTGGGA 3590
```

```
RESULT   3
US-08-131-365B-53
; Sequence 53, Application US/08131365B
; Patent No. 5527690
;  GENERAL INFORMATION:
;    APPLICANT:  Brown, Michael S.
;    APPLICANT:  Briggs, Michael R.
;    APPLICANT:  Wang, Xiaodong
;    APPLICANT:  Goldstein, Joseph L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS RELATING
;    TITLE OF INVENTION:  TO STEROL REGULATORY ELEMENT BINDING
;    TITLE OF INVENTION:  PROTEINS
;    NUMBER OF SEQUENCES:  64
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  U.S.A.
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/131,365B
;      FILING DATE:  01-OCT-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  UTSD:372/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  53:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  5197 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "DNA"
US-08-131-365B-53


  Query Match             3.0%;  Score 114.4;  DB 1;  Length 5197;
  Best Local Similarity   52.3%;  Pred. No. 5.8e-23;
  Matches  315;  Conservative    0;  Mismatches  266;  Indels   21;  Gaps    2;

Qy   2909 cctctgcgatgagtcgagtgtggagctccaagagtgcttgactgtcaaccggatcaccga 2968
          ||  |||||   | | ||||   | ||        || | |  ||||   ||  ||||
Db   2946 CCATTGCGAGAGGGCCAGTGGCCACCTATGGAGCAGCCTCAACGTCAGTGGGGGCACCTC 3005

Qy   2969 cgccaagggtataaagctgcttttccagttgcttacctgcgattggctgctcgaaactag 3028
          | |   |   | || |   |  ||||| |||| ||||| ||   ||| ||       |
Db   3006 TGACCCTGCCCTCAACCACGTGGTCCAGCTGCTCACCTGTGACCTGCTACTGTCGCTACG 3065

Qy   3029 gactgctctgtgggaactggaacacatgaatatggaggacgatgg---cttctaccaagt 3085
          ||| || || ||| ||    |   ||     |   ||| |  ||      ||||||
Db   3066 GACAGCGCTCTGGCAAAAACAGGCCAGTGCCAGCCAGGCTGTGGGGGAGACCTACCACGC 3125

Qy   3086 gccaggtgaagtgctcgagaagttccagaccgatttgaactcgttgcgcaacattgtgga 3145
          | |||| |  |  || | |  |||||    ||  ||  |   ||||||   | |   |
Db   3126 GTCAGGCGCTGAACTGGCGGGCTTCCAACGGGACCTGGGCAGCCTGCGCAGGCTGGCACA 3185

Qy   3146 gaatataccgaacgcccaatcgcgcatatatttgtacgaggcagtttgtcgcctgatggc 3205
          |   | |     ||  |        | |   || | || ||       ||||||||||
Db   3186 CAGCTTCCGCCCAGCATACCGCAAGGTGTTCCTGCATGAAGCCACCGTGCGCCTGATGGC 3245

Qy   3206 tggagcctcaccgtgtccaacgcaacagctcttggacaggagtctgcgatcacgcaacgc 3265
          ||||  |   ||    |  || || |||||  ||||    || |||||    |||| | |
Db   3246 AGGAGGCAGCCCCACCCGCACCCACCAGCTGCTGGAACACAGCCTGCGGCGGCGCACCAC 3305

Qy   3266 ccactcgtccatcttctgcggcagcaaggatcggcggcagcagaacttcgtgggcggaga 3325
          ||                  ||||   | |   |  ||  | |   |  |  |||  |
Db   3306 GCA-----------------GAGCACCAAGCACGGAGAGGTGGATGCCTGGCCCGGCCA 3347

Qy   3326 gcgggaacgggcttcggccatgtacgtggcctgcaagtatctcccgcctgcgctgctcag 3385
          |||  || |||||  | |||||     ||||||||    | ||  || |   |  | |||
Db   3348 GCGAGAGCGGGCCACCGCCATCCTGCTGGCCTGCCGCCACCTGCCCCTCTCCTTCCTCTC 3407

Qy   3386 ctccccgggtgaacgtgctggcatgttagccgaggcggccaagaccctggagaaggtggg 3445
          |||||||||  | ||  || |   || | ||||| ||  |||   ||||||||||||||||
Db   3408 CTCCCCGGGCCAGCGGGCAGTGCTGCTGGCCGAAGCTGCCCGCACCCTGGAGAAGGTGGG 3467

Qy   3446 cgacaagcgaaagctcaaggagtgctaccagctgatgaagtcgctgggcaacggcattgg 3505
          ||||  |||       ||| || ||| | ||| ||||      ||||||    |||| ||
Db   3468 CGACCGGCGCTCCTGCAACGACTGCCAGCAGATGATTGTTAAGCTGGGTGGTGGCACTGC 3527

Qy   3506 ca 3507
          ||
Db   3528 CA 3529


RESULT   4
US-08-668-123-53
; Sequence 53, Application US/08668123
; Patent No. 5891631
;  GENERAL INFORMATION:
;    APPLICANT:  Brown, Michael S.
;    APPLICANT:  Briggs, Michael R.
;    APPLICANT:  Wang, Xiaodong
;    APPLICANT:  Goldstein, Joseph L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS RELATING
;    TITLE OF INVENTION:  TO STEROL REGULATORY ELEMENT BINDING
;    TITLE OF INVENTION:  PROTEINS
;    NUMBER OF SEQUENCES:  64
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  U.S.A.
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/668,123
;      FILING DATE:  14-JUN-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/131,365
;      FILING DATE:  01-OCT-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  UTSD:372/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  53:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5197 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "DNA"
US-08-668-123-53


  Query Match             3.0%;  Score 114.4;  DB 3;  Length 5197;
  Best Local Similarity   52.3%;  Pred. No. 5.8e-23;
  Matches  315;  Conservative    0;  Mismatches  266;  Indels   21;  Gaps    2;

Qy   2909 cctctgcgatgagtcgagtgtggagctccaagagtgcttgactgtcaaccggatcaccga 2968
          ||  |||||   | | ||||   | ||        || | |  ||||   ||  ||||
Db   2946 CCATTGCGAGAGGGCCAGTGGCCACCTATGGAGCAGCCTCAACGTCAGTGGGGGCACCTC 3005

Qy   2969 cgccaagggtataaagctgcttttccagttgcttacctgcgattggctgctcgaaactag 3028
          | |   |   | || |   |  ||||| |||| ||||| ||   ||| ||       |
Db   3006 TGACCCTGCCCTCAACCACGTGGTCCAGCTGCTCACCTGTGACCTGCTACTGTCGCTACG 3065

Qy   3029 gactgctctgtgggaactggaacacatgaatatggaggacgatgg---cttctaccaagt 3085
          ||| || || ||| ||    |   ||     |   ||| |  ||      ||||||
Db   3066 GACAGCGCTCTGGCAAAAACAGGCCAGTGCCAGCCAGGCTGTGGGGGAGACCTACCACGC 3125

Qy   3086 gccaggtgaagtgctcgagaagttccagaccgatttgaactcgttgcgcaacattgtgga 3145
          | |||| |  |  || | |  |||||    ||  ||  |   ||||||   | |   |
Db   3126 GTCAGGCGCTGAACTGGCGGGCTTCCAACGGGACCTGGGCAGCCTGCGCAGGCTGGCACA 3185

Qy   3146 gaatataccgaacgcccaatcgcgcatatatttgtacgaggcagtttgtcgcctgatggc 3205
          |   | |     ||  |        | |   || | || ||       ||||||||||
Db   3186 CAGCTTCCGCCCAGCATACCGCAAGGTGTTCCTGCATGAAGCCACCGTGCGCCTGATGGC 3245

Qy   3206 tggagcctcaccgtgtccaacgcaacagctcttggacaggagtctgcgatcacgcaacgc 3265
          ||||  |   ||    |  || || |||||  ||||    || |||||    |||| | |
Db   3246 AGGAGGCAGCCCCACCCGCACCCACCAGCTGCTGGAACACAGCCTGCGGCGGCGCACCAC 3305

Qy   3266 ccactcgtccatcttctgcggcagcaaggatcggcggcagcagaacttcgtgggcggaga 3325
          ||                  ||||   | |   |  ||  | |   |  |  |||  |
Db   3306 GCA-----------------GAGCACCAAGCACGGAGAGGTGGATGCCTGGCCCGGCCA 3347

Qy   3326 gcgggaacgggcttcggccatgtacgtggcctgcaagtatctcccgcctgcgctgctcag 3385
          |||  || |||||  | |||||     ||||||||    | ||  || |   |  | |||
Db   3348 GCGAGAGCGGGCCACCGCCATCCTGCTGGCCTGCCGCCACCTGCCCCTCTCCTTCCTCTC 3407

Qy   3386 ctccccgggtgaacgtgctggcatgttagccgaggcggccaagaccctggagaaggtggg 3445
          |||||||||  | ||  || |   || | ||||| ||  |||   ||||||||||||||||
```

```
Db   3408 CTCCCCGGGCCAGCGGGCAGTGCTGCTGGCCGAAGCTGCCCGCACCCTGGAGAAGGTGGG 3467

Qy   3446 cgacaagcgaaagctcaaggagtgctaccagctgatgaagtcgctgggcaacggcattgg 3505
          ||||  |||      ||| || ||| | ||| ||||      ||||||    |||| ||
Db   3468 CGACCGGCGCTCCTGCAACGACTGCCAGCAGATGATTGTTAAGCTGGGTGGTGGCACTGC 3527

Qy   3506 ca 3507
          ||
Db   3528 CA 3529


RESULT   5
US-08-469-802B-3
; Sequence 3, Application US/08469802B
; Patent No. 5741645
;  GENERAL INFORMATION:
;    APPLICANT:  Orr, Harry T.
;    APPLICANT:  Ranum, Laura P.W.
;    APPLICANT:  Chung, Ming-yi
;    APPLICANT:  Zoghbi, Huda Y.
;    TITLE OF INVENTION:  Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5741645
;    TITLE OF INVENTION:  Type 1 and Method for Diagnosis
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Mueting, Raasch, Gebhardt & Schwappach, P.A.
;      STREET:  119 No. 5741645th Fourth Street, Suite 203
;      CITY:  Minneapolis
;      STATE:  MN
;      COUNTRY:  USA
;      ZIP:  55401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,802B
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mueting, Ann M.
;      REGISTRATION NUMBER:  33,977
;      REFERENCE/DOCKET NUMBER:  110.00030101
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  612-305-1217
;      TELEFAX:  612-305-1225
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  234 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-469-802B-3


  Query Match             1.3%;  Score 48.4;  DB 2;  Length 234;
  Best Local Similarity   51.9%;  Pred. No. 0.00013;
  Matches  109;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy    314 ggacctggcgccgacgcagatgtacaacatgctgctggacgagcctcgaacgcataccca 373
          | | | |  || |  ||||  | |  |   || || | |  |||  |    |||    ||
Db     24 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 83

Qy    374 gcagacgcagtccgtggatcagcagccgcaatccgtcgagcaacagccgcacgtgaaaag 433
          ||||  ||||     | | ||||||| |||        |||| |||| |||   | |
Db     84 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 143

Qy    434 cgagcactcttcgccagtgcacatcaaggaggaactgcatcagcagcaacaacagtcgcc 493
            ||||      ||    |||     || || | | ||| ||||||| || |||  ||
Db    144 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 203

Qy    494 gcttctcgtctacaaaccagatcccctcat 523
          ||  | | ||  ||   |       |||||
Db    204 GCAGCACCTCAGCAGGGCTCCGGGGCTCAT 233


RESULT   6
US-08-267-803B-3
; Sequence 3, Application US/08267803B
; Patent No. 5834183
;  GENERAL INFORMATION:
;    APPLICANT:  Orr, Harry T.
;    APPLICANT:  Ranum, Laura P.W.
;    APPLICANT:  Chung, Ming-yi
;    APPLICANT:  Zoghbi, Huda Y.
;    TITLE OF INVENTION:  Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5834183
;    TITLE OF INVENTION:  Type 1 and Method for Diagnosis
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Mueting, Raasch, Gebhardt & Schwappach, P.A.
;      STREET:  P.O. Box 581415
;      CITY:  Minneapolis
;      STATE:  MN
;      COUNTRY:  USA
;      ZIP:  55458-1415
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/267,803B
;      FILING DATE:  28-JUN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McCormack, Myra H.
;      REGISTRATION NUMBER:  36,602
;      REFERENCE/DOCKET NUMBER:  110.00030120
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  612-305-1217
;      TELEFAX:  612-305-1228
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  234 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-267-803B-3


  Query Match             1.3%;  Score 48.4;  DB 3;  Length 234;
  Best Local Similarity   51.9%;  Pred. No. 0.00013;
  Matches  109;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy    314 ggacctggcgccgacgcagatgtacaacatgctgctggacgagcctcgaacgcataccca 373
          | | | |  || |  ||||  | |  |   || || | |  |||  |    |||    ||
Db     24 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 83

Qy    374 gcagacgcagtccgtggatcagcagccgcaatccgtcgagcaacagccgcacgtgaaaag 433
          ||||  ||||     | | ||||||| |||        |||| |||| |||   | |
Db     84 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 143

Qy    434 cgagcactcttcgccagtgcacatcaaggaggaactgcatcagcagcaacaacagtcgcc 493
            ||||      ||    |||     || || | | ||| ||||||| || |||  ||
Db    144 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 203

Qy    494 gcttctcgtctacaaaccagatcccctcat 523
          ||  | | ||  ||   |       |||||
Db    204 GCAGCACCTCAGCAGGGCTCCGGGGCTCAT 233


RESULT   7
```

```
US-08-209-747-1/c
; Sequence 1, Application US/08209747
; Patent No. 5733771
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Colgin, Mark
;    TITLE OF INVENTION:  cDNAs Encoding Minor Ampullate Spider
;    TITLE OF INVENTION:  Silk Proteins
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/209,747
;      FILING DATE:  14-MAR-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Nephila clavipes
;      TISSUE TYPE:  minor ampullate gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  183..2675
;      OTHER INFORMATION:  /product= "N. clavipes minor
;      OTHER INFORMATION:  ampullate silk protein"
US-08-209-747-1


  Query Match             1.3%;  Score 47.2;  DB 2;  Length 2793;
  Best Local Similarity   50.0%;  Pred. No. 0.0015;
  Matches  118;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy     268 acgccgacctgcactcggacatgctggacatcatcctcaacgacatggacctggcgccga 327
           ||| | |||| ||| || || | | ||  ||| |  | | | ||| ||| |
Db    1927 ACGACCACCTCTACTGTAACCCGCAGCGCCTCCAGCTCCTGCTCCAGCAGCTGCCGCAGC 1868

Qy     328 cgcagatgtacaacatgctgctggacgagcctcgaacgcatacccagcagacgcagtccg 387
             ||| | |    |  ||  |    | | |||      | || ||  |||   | |  |
Db    1867 ACCAGCTCTTGCTCCGGCACCGTATCCACCCTGACCACCGTAGCCTCCAGCTCCTGCTCC 1808

Qy     388 tggatcagcagccgcaatccgtcgagcaacagccgcacgtgaaaagcgagcactcttcgc 447
           || | | ||||| ||      || |||| | || |||  |  |   | |||||      |
Db    1807 TGCACCTGCAGCGGCTCCAGCTCCAGCACCTGCAGCAGCTCCAGCTCCAGCACCAGCACC 1748

Qy     448 cagtgcacatcaaggaggaactgcatcagcagcaacaacagtcgccgcttctcgtc 503
               | | | || |    |||| | ||| || || ||| |||||  |   ||
Db    1747 TCTACCGTAACCAGCACCTCCTGCTCCTGCACCAGCAGCAGCAGCCGCACCAGCTC 1692


RESULT   8
```

```
US-08-458-298-1/c
; Sequence 1, Application US/08458298
; Patent No. 5756677
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Colgin, Mark
;    TITLE OF INVENTION:  cDNAs Encoding Minor Ampullate Spider
;    TITLE OF INVENTION:  Silk Proteins
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,298
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/209,747
;      FILING DATE:  14-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Nephila clavipes
;      TISSUE TYPE:  minor ampullate gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  183..2675
;      OTHER INFORMATION:  /product= "N. clavipes minor
;      OTHER INFORMATION:  ampullate silk protein"
US-08-458-298-1


  Query Match             1.3%;  Score 47.2;  DB 2;  Length 2793;
  Best Local Similarity   50.0%;  Pred. No. 0.0015;
  Matches  118;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy     268 acgccgacctgcactcggacatgctggacatcatcctcaacgacatggacctggcgccga 327
           ||| | |||| ||| || || | | ||  ||| |  | | | ||| ||| |
Db    1927 ACGACCACCTCTACTGTAACCCGCAGCGCCTCCAGCTCCTGCTCCAGCAGCTGCCGCAGC 1868

Qy     328 cgcagatgtacaacatgctgctggacgagcctcgaacgcatacccagcagacgcagtccg 387
             ||| | |    |  ||  |    | | |||      | || ||  |||   | |  |
Db    1867 ACCAGCTCTTGCTCCGGCACCGTATCCACCCTGACCACCGTAGCCTCCAGCTCCTGCTCC 1808

Qy     388 tggatcagcagccgcaatccgtcgagcaacagccgcacgtgaaaagcgagcactcttcgc 447
           || | | ||||| ||      || |||| | || |||  |  |   | |||||      |
Db    1807 TGCACCTGCAGCGGCTCCAGCTCCAGCACCTGCAGCAGCTCCAGCTCCAGCACCAGCACC 1748

Qy     448 cagtgcacatcaaggaggaactgcatcagcagcaacaacagtcgccgcttctcgtc 503
               | | | || |    |||| | ||| || || ||| |||||  |   ||
Db    1747 TCTACCGTAACCAGCACCTCCTGCTCCTGCACCAGCAGCAGCAGCCGCACCAGCTC 1692
```

```
RESULT   9
US-08-469-802B-2
; Sequence 2, Application US/08469802B
; Patent No. 5741645
;  GENERAL INFORMATION:
;    APPLICANT:  Orr, Harry T.
;    APPLICANT:  Ranum, Laura P.W.
;    APPLICANT:  Chung, Ming-yi
;    APPLICANT:  Zoghbi, Huda Y.
;    TITLE OF INVENTION:  Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5741645
;    TITLE OF INVENTION:  Type 1 and Method for Diagnosis
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Mueting, Raasch, Gebhardt & Schwappach, P.A.
;      STREET:  119 No. 5741645th Fourth Street, Suite 203
;      CITY:  Minneapolis
;      STATE:  MN
;      COUNTRY:  USA
;      ZIP:  55401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,802B
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mueting, Ann M.
;      REGISTRATION NUMBER:  33,977
;      REFERENCE/DOCKET NUMBER:  110.00030101
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  612-305-1217
;      TELEFAX:  612-305-1225
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  195 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-469-802B-2


  Query Match             1.2%;  Score 45.8;  DB 2;  Length 195;
  Best Local Similarity   52.3%;  Pred. No. 0.00066;
  Matches  101;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy    329 gcagatgtacaacatgctgctggacgagcctcgaacgcatacccagcagacgcagtccgt 388
          ||||  | |  |   || || | |  |||  |    |||    ||||||  ||||
Db      3 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 62

Qy    389 ggatcagcagccgcaatccgtcgagcaacagccgcacgtgaaaagcgagcactcttcgcc 448
          | | |||||||| |||        |||| |||| |||   | |     ||||      ||
Db     63 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 122

Qy    449 agtgcacatcaaggaggaactgcatcagcagcaacaacagtcgccgcttctcgtctacaa 508
              |||     || || | | ||| |||||||| || |||  || ||  |||  |
Db    123 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCACCTCAGCAGGGC 182

Qy    509 accagatcccctc 521
           || |  | | ||
Db    183 TCCGGGGCTCATC 195


RESULT  10
US-08-267-803B-2
; Sequence 2, Application US/08267803B
; Patent No. 5834183
;  GENERAL INFORMATION:
;    APPLICANT:  Orr, Harry T.
;    APPLICANT:  Ranum, Laura P.W.
;    APPLICANT:  Chung, Ming-yi
;    APPLICANT:  Zoghbi, Huda Y.
;    TITLE OF INVENTION:  Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5834183
;    TITLE OF INVENTION:  Type 1 and Method for Diagnosis
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Mueting, Raasch, Gebhardt & Schwappach, P.A.
;      STREET:  P.O. Box 581415
;      CITY:  Minneapolis
;      STATE:  MN
;      COUNTRY:  USA
;      ZIP:  55458-1415
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/267,803B
;      FILING DATE:  28-JUN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McCormack, Myra H.
;      REGISTRATION NUMBER:  36,602
;      REFERENCE/DOCKET NUMBER:  110.00030120
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  612-305-1217
;      TELEFAX:  612-305-1228
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  195 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-267-803B-2


  Query Match             1.2%;  Score 45.8;  DB 3;  Length 195;
  Best Local Similarity   52.3%;  Pred. No. 0.00066;
  Matches  101;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy    329 gcagatgtacaacatgctgctggacgagcctcgaacgcatacccagcagacgcagtccgt 388
          ||||  | |  |   || || | |  |||  |    |||    ||||||  ||||
Db      3 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 62

Qy    389 ggatcagcagccgcaatccgtcgagcaacagccgcacgtgaaaagcgagcactcttcgcc 448
          | | |||||||| |||        |||| |||| |||   | |     ||||      ||
Db     63 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 122

Qy    449 agtgcacatcaaggaggaactgcatcagcagcaacaacagtcgccgcttctcgtctacaa 508
              |||     || || | | ||| |||||||| || |||  || ||  |||  |
Db    123 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCACCTCAGCAGGGC 182

Qy    509 accagatcccctc 521
           || |  | | ||
Db    183 TCCGGGGCTCATC 195


RESULT  11
US-08-728-323A-1
; Sequence 1, Application US/08728323A
; Patent No. 5948676
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
```

```
;    TITLE OF INVENTION:  Immediate Early Protein From Kaposi's
;    TITLE OF INVENTION:  Sarcoma-Associated Herpesvirus, DNA
;    TITLE OF INVENTION:  Encoding Same And Uses Thereof
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/728,323A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  0575/52268/JPW/MSC/SKS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-278-0400
;      TELEFAX:  212-391-0525
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3489 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..3489
US-08-728-323A-1


  Query Match             1.2%;  Score 44.8;  DB 4;  Length 3489;
  Best Local Similarity   53.4%;  Pred. No. 0.0089;
  Matches   94;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy    321 gcgccgacgcagatgtacaacatgctgctggacgagcctcgaacgcatacccagcagacg 380
          | || |  |||| | |  |   || || ||| ||||  |    |||     |||||  |
Db   2131 GAGCAGGAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAG 2190

Qy    381 cagtccgtggatcagcagccgcaatccgtcgagcaacagccgcacgtgaaaagcgagcac 440
          |||   |  || |||||||| |||    |   |||| |||| | | | | |     ||||
Db   2191 CAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAG 2250

Qy    441 tcttcgccagtgcacatcaaggaggaactgcatcagcagcaacaacagtcgccgct 496
            |   | | |||     ||||| |   ||| ||| || | ||  ||  |  | ||
Db   2251 GATGAACAGGAGCAGCAGGAGGAGCAGGAGCAGCAGGAGGAGCAGGAGCAGGAGTT 2306


RESULT  12
US-08-770-379-20/c
; Sequence 20, Application US/08770379
; Patent No. 5849564
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
;    TITLE OF INVENTION:  HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/770,379
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  52342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32207 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-770-379-20


  Query Match             1.2%;  Score 44.8;  DB 3;  Length 32207;
  Best Local Similarity   53.4%;  Pred. No. 0.039;
  Matches   94;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy    321 gcgccgacgcagatgtacaacatgctgctggacgagcctcgaacgcatacccagcagacg 380
          | || |  |||| | |  |   || || ||| ||||  |    |||     |||||  |
Db  19866 GAGCAGGAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAG 19807

Qy    381 cagtccgtggatcagcagccgcaatccgtcgagcaacagccgcacgtgaaaagcgagcac 440
          |||   |  || |||||||| |||    |   |||| |||| | | | | |     ||||
Db  19806 CAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAG 19747

Qy    441 tcttcgccagtgcacatcaaggaggaactgcatcagcagcaacaacagtcgccgct 496
            |   | | |||     ||||| |   ||| ||| || | ||  ||  |  | ||
Db  19746 GATGAACAGGAGCAGCAGGAGGAGCAGGAGCAGCAGGAGGAGCAGGAGCAGGAGTT 19691


RESULT  13
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-Fls
US-08-232-463-14


  Query Match             1.2%;  Score 44;  DB 1;  Length 7218;
  Best Local Similarity   4.3%;  Pred. No. 0.025;
  Matches   17;  Conservative  213;  Mismatches  168;  Indels    0;  Gaps    0;

Qy     904 caccgacaagcacaggtagtcgggccagcaaggtgcgagtggcaccactggctccgtcac 963
           |:  :: ::: : ::: ::  :::  :: :::: : ::: :: :  :  ::    :  :
Db    1435 CRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1376

Qy     964 ctgccgctatggaagtccagggcaaggtacctatcaaccgggttcaacccaaggtgaagg 1023
             :  : : :::::   :::: :::: :   :  ::  :::   ::   :::: :::::
Db    1375 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1316

Qy    1024 aagtaaagcgctcggcccacaacgccatcgagcggcgctatcgcacctcaatcaacgaca 1083
           ::: :::: :   ::   : :: :  :  ::: :: :  :  : :    ::  :: :: :
Db    1315 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1256

Qy    1084 agattaacgagttgaagaacttggtagtgggagagcaggccaagctgaacaagtccgcag 1143
           :::  :: :::  ::::::   :: :: :::::::: :::  :::  ::: :::   : ::
Db    1255 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1196

Qy    1144 tgttgcggaaatccatagacaagattcgggatctgcaacgccagaatcacgatctgaagg 1203
           :  : :::::   : ::: ::::  ::::   : :: :  ::::  : ::   :::::
Db    1195 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1136

Qy    1204 cagagttgcagcgcctgcagagggagctaatggcacgcgacggctccaaggtgaaggatt 1263
           ::::  : :: :   : :::::::  :: :: : : :: ::    :::: ::::::
Db    1135 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1076

Qy    1264 tacttcagctgggcactcggcctggtagagcatccaag 1301
           :    ::    |||  |  ||| |    ||| |||||
Db    1075 RRRRRRRRRATCGCAAGCTCCCTCGACCTGCAGCCAAG 1038


RESULT  14
5273901-6
;Patent No. 5273901
;    APPLICANT: JACOBSON, JAMES W.;STRAUSBERG, ROBERT L.;WILSON,
;SUSAN D.;POPE, SHARON H.;STRAUSBERG, SUSAN L.;RUFF,MICHAEL D.;
;AUGUSTINE, PATRICIA C.;DANFORTH, HARRY D.
;    TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS
;SPOROZOITE 21.5 KB ANTIGEN, AC-6B
;    NUMBER OF SEQUENCES: 11
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/581,693
;      FILING DATE: 12-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 215,162
;      FILING DATE: 05-JUL-1988
;      APPLICATION NUMBER: 746,520
;      FILING DATE: 19-JUN-1985
;      APPLICATION NUMBER: 627,811
;      FILING DATE: 05-JUL-1984
;SEQ ID NO:6:
;      LENGTH:543
5273901-6


  Query Match             1.1%;  Score 42.2;  DB 7;  Length 543;
  Best Local Similarity   49.8%;  Pred. No. 0.015;
  Matches  107;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy     274 acctgcactcggacatgctggacatcatcctcaacgacatggacctggcgccgacgcaga 333
           | | |||   | |   || | | |    |   || |  || | |||  ||    || | |
Db     153 aacagcagctgcaacagcagcagctacagcaacaccagctgcagctgcagcaccagctgc 212

Qy     334 tgtacaacatgctgctggacgagcctcgaacgcatacccagcagacgcagtccgtggatc 393
           | |  ||  ||||| | |  |||  |     ||    | ||||  |||     || | |
Db     213 agcagcaccagctgcagcagcagcagcagcaacaggagctgcagcagcaccagctgcagc 272

Qy     394 agcagccgcaatccgtcgagcaacagccgcacgtgaaaagcgagcactcttcgccagtgc 453
           ||||||||    |   ||||  ||| |||   | |  |  |||||     |  |  |
Db     273 gacagccgcagcaggagcagcaggagcagcaaaggcagcggaagcacgaacagaaaccga 332

Qy     454 acatcaaggaggaactgcatcagcagcaacaacag 488
           |     || |||| | ||| ||| ||||    | ||
Db     333 aggagcagcaggagcagcagaagcagagaaag 367


RESULT  15
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```

```
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             1.1%;  Score 41.6;  DB 1;  Length 7218;
  Best Local Similarity   3.1%;  Pred. No. 0.12;
  Matches   11;  Conservative  197;  Mismatches  146;  Indels    0;  Gaps    0;

Qy     554 gccgacgggccttttgaaggccgcccaaccaacagccaccatacatcacatggacgccca 613
           | ||| ||  ||| ||   :: ::: :: : :: :: : : :: : :   : :::
Db    1049 GTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1108

Qy     614 gcggatgccgccgaacacggcggtgtatcccccatctctgggcagtagctttgtctacca 673
            :   : :: ::   : :  :  : : :::::: :::::   :  :  :::: ::: ::
Db    1109 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1168

Qy     674 gtccatgtccccgcccacgtcgccggtggagtctgcgaaccagaatgtcaatgtcatgca 733
            ::: : ::::: ::: : :: ::  :    ::: :   ::    : ::  : :: : :
Db    1169 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1228

Qy     734 gcccgttgctgcaactcctgctcccgcttctgctcctttgccccagcagtcgtatccgca 793
            ::: :: :: :  ::::: ::::: ::::: :::::::: :::: :  :: : ::: :
Db    1229 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1288

Qy     794 acccttcattacgtacaactctaaggccggaatgacttccgatgaagccatgtacttgct 853
            :::::: :: : : :  ::::    ::    :  ::::: :    :: : : ::: ::
Db    1289 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1348

Qy     854 cttgcagcccacggtagccagtccaaccccatctccacctgtggctccaccacc 907
           ::: :  ::: :  :  ::  :::  :::: ::::: ::: :  :::: :: ::
Db    1349 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1402


Search completed: September  4, 2000, 00:36:23
Job time: 20353 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2000, 00:25:16 ; Search time 4457.97 Seconds
(without alignments)
3727.309 Million cell updates/sec

Title: US-09-332-522B-7
Perfect score: 3768
Sequence: 1 gtttattaagctgcaaatat.....aatattttaacagatttaaa 3768

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: em_est35:*
83: em_est36:*
84: em_est37:*
85: gb_est48:*
86: gb_est49:*
87: gb_est50:*
88: gb_est51:*
89: gb_est52:*
90: gb_est53:*
91: gb_est54:*
92: gb_est55:*
93: gb_gss1:*
94: gb_gss2:*
95: gb_gss3:*
96: gb_gss4:*
97: em_gss1:*
98: em_gss2:*
99: em_gss3:*
100: em_gss4:*
101: gb_gss5:*
102: gb_gss6:*
103: gb_gss7:*
104: gb_gss8:*
105: gb_gss9:*
106: em_gss5:*
107: em_gss6:*
108: em_gss7:*
109: em_gss8:*
110: em_gss9:*
111: em_gss10:*
112: em_gss11:*
113: gb_gss10:*
114: gb_gss11:*
115: em_gss12:*
116: gb_gss12:*

```
            117:  gb_gss13:*
            118:  gb_gss14:*
            119:  gb_gss15:*
            120:  gb_gss16:*
            121:  gb_gss17:*
            122:  gb_gss18:*
            123:  gb_gss19:*
            124:  em_gss13:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		638	16.9	639	35	AI114160	AI114160 GH10860.5
2		631	16.7	631	35	AI107040	AI107040 GH06327.5
3		589.4	15.6	598	28	AA567826	AA567826 HL01847.5
4		534.4	14.2	552	40	AI533148	AI533148 SD04874.5
5		519.4	13.8	521	40	AI533491	AI533491 SD05333.5
6		461	12.2	487	22	AA201358	AA201358 LD04332.5
7		175	4.6	282	37	AI297927	AI297927 LP12374.5
8	c	118.4	3.1	1101	122	CNS0106C	AL098574 Drosophil
9		100.2	2.7	352	47	AL043416	AL043416 DKFZp434I
10		91.8	2.4	506	80	AW701427	AW701427 uq86a04.y
11		89.6	2.4	574	47	AJ396343	AJ396343
12		80.6	2.1	554	47	AL040138	AL040138 DKFZp434D
13	c	73	1.9	443	31	AA806181	AA806181 oc25e08.s
14		70.8	1.9	524	39	AI386259	AI386259 mm45e05.y
15		70.4	1.9	412	31	AA823111	AA823111 vw40g09.r
16	c	70	1.9	619	123	FR0039256	AL126754 Fugu rubr
17	c	70	1.9	791	35	AI132428	AI132428 ue25b02.x
18		69	1.8	443	29	AA673593	AA673593 vn47a05.r
19		67	1.8	226	28	AA597195	AA597195 vo24c04.r
20		66.8	1.8	529	44	AI786503	AI786503 uf15g04.y
21		64.6	1.7	487	73	AW502177	AW502177 UI-HF-BR0
22	c	63	1.7	543	71	AW322518	AW322518 uo76h03.x
23		62.4	1.7	482	42	AI607243	AI607243 vo64d01.y
24	c	62.4	1.7	520	34	AI013042	AI013042 EST207493
25	c	62.4	1.7	724	38	AI326423	AI326423 mm45e05.x
26	c	62	1.6	565	69	AW227538	AW227538 up10d09.x
27	c	61.4	1.6	577	72	AW411854	AW411854 uq44f02.x
28	c	60.4	1.6	543	41	AI551224	AI551224 vo64d01.x
29	c	59.2	1.6	613	69	AW214475	AW214475 uo49f02.x
30	c	58	1.5	300	69	AW205250	AW205250 UI-H-BI1-
31	c	57	1.5	504	69	AW228716	AW228716 up19f02.x
32	c	56.2	1.5	542	69	AW227575	AW227575 up11b02.x
33		56	1.5	495	79	AW661326	AW661326 833004D05
34	c	56	1.5	536	80	AW701332	AW701332 uq86a04.x
35		55.6	1.5	410	42	AI644706	AI644706 vb01d11.y
36	c	55.4	1.5	221	45	AI875765	AI875765 uj21d05.x
37	c	55.4	1.5	824	36	AI201526	AI201526 qs74e07.x
38	c	55	1.5	542	69	AW226536	AW226536 up08g12.x
39		53.8	1.4	306	28	AA620226	AA620226 vo64d01.r
40		53.8	1.4	542	73	AW504363	AW504363 UI-HF-BN0
41		52.8	1.4	496	86	H49228	H49228 yq19h01.r1
42		52.6	1.4	485	22	AA158187	AA158187 zo55d11.r
43		51.8	1.4	598	123	FR0039269	AL126767 Fugu rubr
44		51.2	1.4	261	73	AW483186	AW483186 50631 MAR
45	c	50.6	1.3	506	69	AW228679	AW228679 up19b08.x

ALIGNMENTS

```
RESULT   1
AI114160
LOCUS       AI114160      639 bp    mRNA            EST       18-NOV-1998
DEFINITION  GH10860.5prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH10860 5prime similar to U38238: HLH106
            FBgn0015234 PID:g1079656 SPTREMBL:Q24146, mRNA sequence.
ACCESSION   AI114160
VERSION     AI114160.1  GI:3514963
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 639)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3189380.
            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 108  row: E  column: 12
            High quality sequence stop: 510.
FEATURES             Location/Qualifiers
     source          1. .639
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="GH10860"
                     /clone_lib="GH Drosophila melanogaster head pOT2"
                     /sex="male and female"
                     /dev_stage="adult"
                     /lab_host="DH5 - alpha"
                     /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2.  Plasmid cDNA library."
BASE COUNT      151 a    219 c    148 g    120 t      1 others
ORIGIN

  Query Match            16.9%;  Score 638;  DB 35;  Length 639;
  Best Local Similarity  99.8%;  Pred. No. 2.6e-162;
  Matches  638;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    232 atttgttcaaagcggaggatgtcttcgaaccgttcgacgccgacctgcactcggacatgc 291
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATTTGTTCAAAGCGGAGGATGTCTTCGAACCGTTCGACGCCGACCTGCACTCGGACATGC 60

Qy    292 tggacatcatcctcaacgacatggacctggcgccgacgcagatgtacaacatgctgctgg 351
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGGACATCATCCTCAACGACATGGACCTGGCGCCGACGCAGATGTACAACATGCTGCTGG 120

Qy    352 acgagcctcgaacgcatacccagcagacgcagtccgtggatcagcagccgcaatccgtcg 411
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ACGAGCCTCGAACGCATACCCAGCAGACGCAGTCCGTGGATCAGCAGCCGCAATCCGTCG 180

Qy    412 agcaacagccgcacgtgaaaagcgagcactcttcgccagtgcacatcaaggaggaactgc 471
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AGCAACAGCCGCACGTGAAAAGCGAGCACTCTTCGCCAGTGCACATCAAGGAGGAACTGC 240

Qy    472 atcagcagcaacaacagtcgccgcttctcgtctacaaaccagatcccctcatagccacaa 531
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ATCAGCAGCAACAACAGTCGCCGCTTCTCGTCTACAAACCAGATCCCCTCATAGCCACAA 300

Qy    532 gctacaattgtccccagcaacagccgacgggccttttgaaggccgcccaaccaacagcca 591
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCTACAATTGTCCCCAGCAACAGCCGACGGGCCTTTTGAAGGCCGCCCAACCAACAGCCA 360

Qy    592 ccatacatcacatggacgcccagcggatgccgccgaacacggcggtgtatcccccatctc 651
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CCATACATCACATGGACGCCCAGCGGATGCCGCCGAACACGGCGGTGTATCCCCCATCTC 420

Qy    652 tgggcagtagctttgtctaccagtccatgtccccgcccacgtcgccggtggagtctgcga 711
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    421 TGGGCAGTAGCTTTGTCTACCAGTCCATGTCCCCGCCCACGTCGCCGGTGGAGTCTGCGA 480

Qy    712 accagaatgtcaatgtcatgcagcccgttgctgcaactcctgctcccgcttctgctcctt 771
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ACCAGAATGTCAATGTCATGCAGCCCGTTGCTGCAACTCCTGCTCCCGCTTCTGCTCCTT 540

Qy    772 tgccccagcagtcgtatccgcaacccttcattacgtacaactctaaggccggaatgactt 831
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TGCCCCAGCAGTCGTATCCGCAACCCTTCATTACGTACAACTCTAAGGCCGGAATGACTT 600

Qy    832 ccgatgaagccatgtacttgctcttgcagcccacggtag 870
          |||||||||||||||||| ||||||||||||||||||
Db    601 CCGATGAAGCCATGTACTTNCTCTTGCAGCCCACGGTAG 639


RESULT   2
AI107040
LOCUS       AI107040      631 bp    mRNA            EST       24-NOV-1998
DEFINITION  GH06327.5prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH06327 5prime similar to U38238: HLH106
            FBgn0015234 PID:g1079656 SPTREMBL:Q24146, mRNA sequence.
ACCESSION   AI107040
VERSION     AI107040.1  GI:3475975
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 631)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Apr 3, 1998 this sequence version replaced gi:3018338.
            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 63  row: C  column: 3
            High quality sequence stop: 539.
FEATURES             Location/Qualifiers
     source          1. .631
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="GH06327"
                     /clone_lib="GH Drosophila melanogaster head pOT2"
                     /sex="male and female"
                     /dev_stage="adult"
                     /lab_host="DH5 - alpha"
                     /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2.  Plasmid cDNA library."
BASE COUNT      200 a    180 c    140 g    111 t
ORIGIN


  Query Match             16.7%;  Score 631;  DB 35;  Length 631;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-160;
  Matches  631;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 tttattaagctgcaaatatactcgtgaaaaaaatcaaaacaaccatgaacaacaagtgtt 61
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TTTATTAAGCTGCAAATATACTCGTGAAAAAAATCAAAACAACCATGAACAACAAGTGTT 60

Qy     62 gcaactattactaactagtcgctagtttaaagcaaagtgcgttgacattaaccagttatg 121
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCAACTATTACTAACTAGTCGCTAGTTTAAAGCAAAGTGCGTTGACATTAACCAGTTATG 120

Qy    122 gaaaaacaaaagcacacgtgaactaagaaaacagatagaaggtggtaaagcattcgcaat 181
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GAAAAACAAAAGCACACGTGAACTAAGAAAACAGATAGAAGGTGGTAAAGCATTCGCAAT 180

Qy    182 ggacacgacactgatgaacttaatagacgctccgctggacgagtccatggatttgttcaa 241
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GGACACGACACTGATGAACTTAATAGACGCTCCGCTGGACGAGTCCATGGATTTGTTCAA 240

Qy    242 agcggaggatgtcttcgaaccgttcgacgccgacctgcactcggacatgctggacatcat 301
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGCGGAGGATGTCTTCGAACCGTTCGACGCCGACCTGCACTCGGACATGCTGGACATCAT 300

Qy    302 cctcaacgacatggacctggcgccgacgcagatgtacaacatgctgctggacgagcctcg 361
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCTCAACGACATGGACCTGGCGCCGACGCAGATGTACAACATGCTGCTGGACGAGCCTCG 360

Qy    362 aacgcatacccagcagacgcagtccgtggatcagcagccgcaatccgtcgagcaacagcc 421
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AACGCATACCCAGCAGACGCAGTCCGTGGATCAGCAGCCGCAATCCGTCGAGCAACAGCC 420

Qy    422 gcacgtgaaaagcgagcactcttcgccagtgcacatcaaggaggaactgcatcagcagca 481
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GCACGTGAAAAGCGAGCACTCTTCGCCAGTGCACATCAAGGAGGAACTGCATCAGCAGCA 480

Qy    482 acaacagtcgccgcttctcgtctacaaaccagatcccctcatagccacaagctacaattg 541
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ACAACAGTCGCCGCTTCTCGTCTACAAACCAGATCCCCTCATAGCCACAAGCTACAATTG 540

Qy    542 tccccagcaacagccgacgggccttttgaaggccgcccaaccaacagccaccatacatca 601
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TCCCCAGCAACAGCCGACGGGCCTTTTGAAGGCCGCCCAACCAACAGCCACCATACATCA 600

Qy    602 catggacgcccagcggatgccgccgaacacg 632
          ||||||||||||||||||||||||||||||
Db    601 CATGGACGCCCAGCGGATGCCGCCGAACACG 631


RESULT   3
AA567826
LOCUS       AA567826      598 bp    mRNA            EST       28-NOV-1998
DEFINITION  HL01847.5prime HL Drosophila melanogaster head BlueScript
            Drosophila melanogaster cDNA clone HL01847 5prime similar to
            U38238: HLH106 FBgn0015234 PID:g1079656 SPTREMBL:Q24146, mRNA
            sequence.
ACCESSION   AA567826
VERSION     AA567826.1  GI:2340608
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 598)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 18  row: D  column: 11
            High quality sequence stop: 350.
FEATURES             Location/Qualifiers
     source          1. .598
                     /organism="Drosophila melanogaster"
                     /db_xref="BDGP_EST:BDcln026410"
                     /db_xref="taxon:7227"
                     /clone="HL01847"
                     /clone_lib="HL Drosophila melanogaster head BlueScript"
                     /sex="male and female"
                     /dev_stage="adult"
                     /lab_host="SOLR"
```

```
                /note="Organ: head--brain & sensory organ; Vector:
                BlueScript SK; Site_1: EcoRI; Site_2: XhoI; Constructed
                using Stratagene ZAP-cDNA Synthesis kit. Oligo dT-primed
                and directionally cloned at EcoRI and XhoI in BlueScript
                SK(+/-)"
BASE COUNT      151 a    181 c    157 g    109 t
ORIGIN


  Query Match             15.6%;  Score 589.4;  DB 28;  Length 598;
  Best Local Similarity   99.8%;  Pred. No. 4e-149;
  Matches  590;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    709 cgaaccagaatgtcaatgtcatgcagcccgttgctgcaactcctgctcccgcttctgctc 768
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGAACCAGAATGTCAATGTCATGCAGCCCGTTGCTGCAACTCCTGCTCCCGCTTCTGCTC 60

Qy    769 ctttgccccagcagtcgtatccgcaacccttcattacgtacaactctaaggccggaatga 828
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CTTTGCCCCAGCAGTCGTATCCGCAACCCTTCATTACGTACAACTCTAAGGCCGGAATGA 120

Qy    829 cttccgatgaagccatgtacttgctcttgcagcccacggtagccagtccaaccccatctc 888
          ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    121 CTTCCGATGAAGCCATGTACTTCCTCTTGCAGCCCACGGTAGCCAGTCCAACCCCATCTC 180

Qy    889 cacctgtggctccaccaccgacaagcacaggtagtcgggccagcaaggtgcgagtggcac 948
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CACCTGTGGCTCCACCACCGACAAGCACAGGTAGTCGGGCCAGCAAGGTGCGAGTGGCAC 240

Qy    949 cactggctccgtcacctgccgctatggaagtccagggcaaggtacctatcaaccgggttc 1008
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CACTGGCTCCGTCACCTGCCGCTATGGAAGTCCAGGGCAAGGTACCTATCAACCGGGTTC 300

Qy   1009 aacccaaggtgaaggaagtaaagcgctcggcccacaacgccatcgagcggcgctatcgca 1068
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AACCCAAGGTGAAGGAAGTAAAGCGCTCGGCCCACAACGCCATCGAGCGGCGCTATCGCA 360

Qy   1069 cctcaatcaacgacaagattaacgagttgaagaacttggtagtgggagagcaggccaagc 1128
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CCTCAATCAACGACAAGATTAACGAGTTGAAGAACTTGGTAGTGGGAGAGCAGGCCAAGC 420

Qy   1129 tgaacaagtccgcagtgttgcggaaatccatagacaagattcgggatctgcaacgccaga 1188
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TGAACAAGTCCGCAGTGTTGCGGAAATCCATAGACAAGATTCGGGATCTGCAACGCCAGA 480

Qy   1189 atcacgatctgaaggcagagttgcagcgcctgcagagggagctaatggcacgcgacggct 1248
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATCACGATCTGAAGGCAGAGTTGCAGCGCCTGCAGAGGGAGCTAATGGCACGCGACGGCT 540

Qy   1249 ccaaggtgaaggatttacttcagctgggcactcggcctggtagagcatcca 1299
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CCAAGGTGAAGGATTTACTTCAGCTGGGCACTCGGCCTGGTAGAGCATCCA 591


RESULT   4
AI533148
LOCUS       AI533148      552 bp    mRNA            EST       18-MAR-1999
DEFINITION  SD04874.5prime SD Drosophila melanogaster Schneider L2 cell culture
            pOT2 Drosophila melanogaster cDNA clone SD04874 5prime similar to
            U38238: HLH106 FBgn0015234 PID:g1079656 SPTREMBL:Q24146, mRNA
            sequence.
ACCESSION   AI533148
VERSION     AI533148.1  GI:4447283
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 552)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 48  row: G  column: 2
            High quality sequence stop: 243.
FEATURES             Location/Qualifiers
     source          1. .552
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="SD04874"
                     /clone_lib="SD Drosophila melanogaster Schneider L2 cell
                     culture pOT2"
                     /lab_host="DH5-alpha"
                     /note="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
                     fractionated cDNAs were directly ligated into pOT2.
                     Plasmid cDNA library."
BASE COUNT      128 a    190 c    135 g     99 t
ORIGIN


  Query Match             14.2%;  Score 534.4;  DB 40;  Length 552;
  Best Local Similarity   98.0%;  Pred. No. 3.4e-134;
  Matches  541;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy    519 ctcatagccacaagctacaattgtccccagcaacagccgacgggccttttgaaggccgcc 578
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CTCATAGCCACAAGCTACAATTGTCCCCAGCAACAGCCGACGGGCCTTTTGAAGGCCGCC 60

Qy    579 caaccaacagccaccatacatcacatggacgcccagcggatgccgccgaacacggcggtg 638
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CAACCAACAGCCACCATACATCACATGGACGCCCAGCGGATGCCGCCGAACACGGCGGTG 120

Qy    639 tatccccatctctgggcagtagctttgtctaccagtccatgtccccgcccacgtcgccg 698
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TATCCCCCATCTGTGGGCAGTAGCTTTGTCTACCAGTCCATGTCCCCGCCCACGTCGCCG 180

Qy    699 gtggagtctgcgaaccagaatgtcaatgtcatgcagcccgttgctgcaactcctgctccc 758
          |||||| |||||||||| ||||  ||||||||||||||||||||||||||||||||||||
Db    181 GTGGAGACTGCGAACCAAAATGGGAATGTCATGCAGCCCGTTGCTGCAACTCCTGCTCCC 240

Qy    759 gcttctgctcctttgccccagcagtcgtatccgcaacccttcattacgtacaactctaag 818
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCTTATGCTCCTTTGCCCCAGCAGTCGTATCCGCAACCCTTCATTACGTACAACTCTAAG 300

Qy    819 gccggaatgacttccgatgaagccatgtacttgctcttgcagcccacggtagccagtcca 878
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCCGGAATGACTTCCGATGAAGCCATGTACTTGCTCTTGCAGCCCACGGTAGCCAGTCCA 360

Qy    879 accccatctccacctgtggctccaccaccgacaagcacaggtagtcgggccagcaaggtg 938
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ACCCCATCTACACCTGTGGCTCCACCACCGACAAGCACAGGTAGTCGGGCCAGCAAGGTG 420

Qy    939 cgagtggcaccactggctccgtcacctgccgctatggaagtccagggcaaggtacctatc 998
          |||||||||||||||| |||||||||| |||||||| ||||||||||||||||||||||
Db    421 CGAGTGGCACCACTGGCTGCGTCACCTGCCGGTATGGAAGACCAGGGCAAGGTACCTATC 480

Qy    999 aaccgggttcaacccaaggtgaaggaagtaaagcgctcggcccacaacgccatcgagcgg 1058
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    481 AACCGGGTTCAACCCAAGGTGAAGGAAGTAAAGCGCTCGGCCCACAATGCCATCGAGCGG 540

Qy   1059 cgctatcgcacc 1070
          ||||||||||||
Db    541 CGCTATCGCACC 552


RESULT   5
AI533491
LOCUS       AI533491      521 bp    mRNA            EST       18-MAR-1999
DEFINITION  SD05333.5prime SD Drosophila melanogaster Schneider L2 cell culture
```

```
            pOT2 Drosophila melanogaster cDNA clone SD05333 5prime similar to
            U38238: HLH106 FBgn0015234 PID:g1079656 SPTREMBL:Q24146, mRNA
            sequence.
ACCESSION   AI533491
VERSION     AI533491.1  GI:4447626
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 521)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 53  row: C  column: 9
            High quality sequence stop: 510.
FEATURES             Location/Qualifiers
     source          1. .521
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="SD05333"
                     /clone_lib="SD Drosophila melanogaster Schneider L2 cell
                     culture pOT2"
                     /lab_host="DH5-alpha"
                     /note="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
                     fractionated cDNAs were directly ligated into pOT2.
                     Plasmid cDNA library."
BASE COUNT      115 a    187 c    118 g    101 t
ORIGIN


  Query Match             13.8%;  Score 519.4;  DB 40;  Length 521;
  Best Local Similarity   99.8%;  Pred. No. 4e-130;
  Matches  520;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    496 ttctcgtctacaaaccagatcccctcatagccacaagctacaattgtccccagcaacagc 555
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TTCTCGTCTACAAACCAGATCCCCTCATAGCCACAAGCTACAATTGTCCCCAGCAACAGC 60

Qy    556 cgacgggccttttgaaggccgcccaaccaacagccaccatacatcacatggacgcccagc 615
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CGACGGGCCTTTTGAAGGCCGCCCAACCAACAGCCACCATACATCACATGGACGCCCAGC 120

Qy    616 ggatgccgccgaacacggcggtgtatccccatctctgggcagtagctttgtctaccagt 675
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGATGCCGCCGAACACGGCGGTGTATCCCCCATCTCTGGGCAGTAGCTTTGTCTACCAGT 180

Qy    676 ccatgtccccgcccacgtcgccggtggagtctgcgaaccagaatgtcaatgtcatgcagc 735
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    181 CCATGTCCCCGCCCACGTCGCCGGTGGAGTCTGCGAACCAGAATGTGAATGTCATGCAGC 240

Qy    736 ccgttgctgcaactcctgctcccgcttctgctcctttgccccagcagtcgtatccgcaac 795
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CCGTTGCTGCAACTCCTGCTCCCGCTTCTGCTCCTTTGCCCCAGCAGTCGTATCCGCAAC 300

Qy    796 ccttcattacgtacaactctaaggccggaatgacttccgatgaagccatgtacttgctct 855
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCTTCATTACGTACAACTCTAAGGCCGGAATGACTTCCGATGAAGCCATGTACTTGCTCT 360

Qy    856 tgcagcccacggtagccagtccaaccccatctccacctgtggctccaccaccgacaagca 915
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGCAGCCCACGGTAGCCAGTCCAACCCCATCTCCACCTGTGGCTCCACCACCGACAAGCA 420

Qy    916 caggtagtcgggccagcaaggtgcgagtggcaccactggctccgtcacctgccgctatgg 975
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CAGGTAGTCGGGCCAGCAAGGTGCGAGTGGCACCACTGGCTCCGTCACCTGCCGCTATGG 480

Qy    976 aagtccagggcaaggtacctatcaaccgggttcaacccaag 1016
          |||||||||||||||||||||||||||||||||||||||||
Db    481 AAGTCCAGGGCAAGGTACCTATCAACCGGGTTCAACCCAAG 521


RESULT   6
AA201358
LOCUS       AA201358      487 bp    mRNA            EST       29-NOV-1998
DEFINITION  LD04332.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD04332 5prime, mRNA sequence.
ACCESSION   AA201358
VERSION     AA201358.1  GI:1797151
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 487)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Blast expect value = 2.5E-185 on U38238: Drosophila melanogaster
            transcription factor HLH106 mRNA, complete cds
            Plate: 43  row: C  column: 8
            High quality sequence stop: 485.
FEATURES             Location/Qualifiers
     source          1. .487
                     /organism="Drosophila melanogaster"
                     /db_xref="BDGP_EST:BDcln003739"
                     /db_xref="taxon:7227"
                     /clone="LD04332"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                     EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      161 a    121 c    112 g     93 t
ORIGIN


  Query Match             12.2%;  Score 461;  DB 22;  Length 487;
  Best Local Similarity   99.8%;  Pred. No. 2.8e-114;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy      1 gtttattaagctgcaaatatactcgtgaaaaaaatcaaaacaaccatgaacaacaagtgt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     16 GTTTATTAAGCTGCAAATATACTCGTGAAAAAAATCAAAACAACCATGAACAACAAGTGT 75

Qy     61 tgcaactattactaactagtcgctagtttaaagcaaagtgcgttgacattaaccagttat 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 TGCAACTATTACTAACTAGTCGCTAGTTTAAAGCAAAGTGCGTTGACATTAACCAGTTAT 135

Qy    121 ggaaaaacaaaagcacacgtgaactaagaaaacagatagaaggtggtaaagcattcgcaa 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 GGAAAAACAAAAGCACACGTGAACTAAGAAAACAGATAGAAGGTGGTAAAGCATTCGCAA 195

Qy    181 tggacacgacactgatgaacttaatagacgctccgctggacgagtccatggatttgttca 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 TGGACACGACACTGATGAACTTAATAGACGCTCCGCTGGACGAGTCCATGGATTTGTTCA 255
```

```
Qy    241 aagcggaggatgtcttcgaaccgttcgacgccgacctgcactcggacatgctggacatca 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 AAGCGGAGGATGTCTTCGAACCGTTCGACGCCGACCTGCACTCGGACATGCTGGACATCA 315

Qy    301 tcctcaacgacatggacctggcgccgacgcagatgtacaacatgctgctggacgagcctc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    316 TCCTCAACGACATGGACCTGGCGCCGACGCAGATGTACAACATGCTGCTGGACGAGCCTC 375

Qy    361 gaacgcatacccagcagacgcagtccgtggatcagcagccgcaatccgtcgagcaacagc 420
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    376 GAACGCATACCCAGCAGACGC-GTCCGTGGATCAGCAGCCGCAATCCGTCGAGCAACAGC 434

Qy    421 cgcacgtgaaaagcgagcactcttcgccagtgcacatcaaggaggaactgcat 473
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    435 CGCACGTGAAAAGCGAGCACTCTTCGCCAGTGCACATCAAGGAGGAACTGCAT 487


RESULT   7
AI297927
LOCUS       AI297927      282 bp    mRNA            EST       01-DEC-1998
DEFINITION  LP12374.5prime LP Drosophila melanogaster larval-early pupal pOT2
            Drosophila melanogaster cDNA clone LP12374 5prime, mRNA sequence.
ACCESSION   AI297927
VERSION     AI297927.1  GI:3947334
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 282)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 123  row: G  column: 2
            High quality sequence stop: 207.
FEATURES             Location/Qualifiers
     source          1. .282
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LP12374"
                     /clone_lib="LP Drosophila melanogaster larval-early pupal
                     pOT2"
                     /sex="male and female"
                     /dev_stage="larvae-pupae"
                     /lab_host="DH5-alpha"
                     /note="Organ: whole body; Vector: pOT2; Site_1: EcoRI;
                     Site_2: Xho1; Sized fractionated cDNAs were directly
                     ligated into pOT2. Plasmid cDNA library.
                     "
BASE COUNT      113 a     53 c     55 g     61 t
ORIGIN


  Query Match             4.6%;  Score 175;  DB 37;  Length 282;
  Best Local Similarity   89.0%;  Pred. No. 1.1e-36;
  Matches  211;  Conservative    0;  Mismatches    0;  Indels   26;  Gaps    1;

Qy      1 gtttattaagctgcaaatatactcgtgaaaaaaatcaaaacaaccatgaacaacaagtgt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     46 GTTTATTAAGCTGCAAATATACTCGTGAAAAAAATCAAAACAACCATGAACAACAAGTGT 105

Qy     61 tgcaactattactaactagtcgctagtttaaagcaaagtgcgttgacattaaccagttat 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    106 TGCAACTATTACTAACTAGTCGCTAGTTTAAAGCAAAGTGCGTTGACATTAACCAGTTAT 165

Qy    121 ggaaaaacaaaagcacacgtgaactaagaaaacagatagaaggtggtaaa---------- 170
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    166 GGAAAAACAAAAGCACACGTGAACTAAGAAAACAGATAGAAGGTGGTAAAGCTGCAAGTG 225

Qy    171 ----------------gcattcgcaatggacacgacactgatgaacttaatagacgc 211
                           ||||||||||||||||||||||||||||||||||||||||
Db    226 GCCACTGTGGAAAAAGGCATTCGCAATGGACACGACACTGATGAACTTAATAGACGC 282


RESULT   8
CNS0106C/c
LOCUS       CNS0106C     1101 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN03K02 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098574
VERSION     AL098574.1  GI:5610185
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Buchetcn
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN03K02"
                     /note="end : T7"
BASE COUNT      277 a    237 c     78 g    237 t    272 others
ORIGIN


  Query Match             3.1%;  Score 118.4;  DB 122;  Length 1101;
  Best Local Similarity   54.6%;  Pred. No. 4e-21;
  Matches  173;  Conservative   12;  Mismatches  129;  Indels    3;  Gaps    2;

Qy    169 aagcattcgcaatggacacgacactgatgaacttaatagacgctccgctggacgagtcca 228
          | | ||| | :||||| : || | || ||   ||  | |  :  | | |||  ||||
Db    324 AGGAATTGGGRATGGAGRGGAAAATGTTGTNKTTKNTTGNAAGTNNGTTGGNAGAGTNNN 265

Qy    229 tggatttgttcaaagcggaggatgtcttcgaaccgttcgacgccgacctgcactcggaca 288
          ||| ||||||    | |||||||||| || |||  ||| || |  |   || | | ||| |
Db    264 TGGNTTTGTTNNNCGNGGAGGATGTTTTTGAANNGTTTGAAGTAGNTTTGTANTGGGANA 205

Qy    289 tgctggacatcatcctcaacgacatggacctggcgccgacgcagatgtacaacatgctgc 348
          || |||| || ||  | || || ||||   ||| |  |  | :|||||  || ||| ||
Db    204 TGNTGGANATNATNNTAAANGANATGGGNTTGGGGGGGGGGGGRGATGTGGAANATGNTGN 145

Qy    349 tggacgagcctcgaacgcatacccagca--gacgcagtccgtggatcagcagccgcaatc 406
          |||| | |  | |||:  |     |: |  || | |||  ||||:|:||  |  | |||
Db    144 TGGANGNGNNTNGAAVCGAATANNARGAACGAGGAAGTNNGTGGRTSAGNGGNGGNAATN 85

Qy    407 cgtcgagcaacagccgcacgtgaaaagcgagcac-tcttcgccagtgcacatcaaggagg 465
          || |||  | ||  |   |||| | | | | || :||| | |:    |:||    || :
Db     84 NGTGGAGNNANAGNNGNNNGTGANAGGNGNGNACAYCTTNGACMMGAAAYATANNAGAAR 25
```

```
Qy    466 aactgcatcagcagcaa 482
          ||: | |     |   |
Db     24 AAMAGNANANANAANNA 8


RESULT   9
AL043416
LOCUS       AL043416      352 bp    mRNA            EST       29-FEB-2000
DEFINITION  DKFZp434I0126_r1 434 (synonym: htes3) Homo sapiens cDNA clone
            DKFZp434I0126 5', mRNA sequence.
ACCESSION   AL043416
VERSION     AL043416.1  GI:5422803
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 352)
  AUTHORS   Ansorge,W., Benes,V., Krieger,S., Mewes,H.W., Gassenhuber,J. and
            Wiemann,S.
  TITLE     EST (Ansorge, Benes, et al.)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Ansorge W
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by EMBL (European Molecular Biology Laboratories,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            s1 sequence also available.
            This clone (DKFZp434I0126) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .352
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp434I0126"
                     /clone_lib="434 (synonym: htes3)"
                     /tissue_type="testis"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT       95 a    102 c     92 g     63 t
ORIGIN


  Query Match             2.7%;  Score 100.2;  DB 47;  Length 352;
  Best Local Similarity   69.9%;  Pred. No. 2.5e-16;
  Matches  135;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy   1029 aagcgctcggcccacaacgccatcgagcggcgctatcgcacctcaatcaacgacaagatt 1088
          |||||| | |||||||||||||| |||  |||||| ||| |||| ||||| ||||| ||
Db      2 AAGCGCACAGCCCACAACGCCATTGAGAAGCGCTACCGCTCCTCCATCAATGACAAAATC 61

Qy   1089 aacgagttgaagaacttggtagtgggagagcaggccaagctgaacaagtccgcagtgttg 1148
          |  ||| | ||| |  |||| ||||     |||| |||||||| || || || || |||
Db     62 ATTGAGCTCAAGGATCTGGTGGTGGGCACTGAGGCAAAGCTGAATAAATCTGCTGTCTTG 121

Qy   1149 cggaaatccatagacaagattcgggatctgcaacgccagaatcacgatctgaaggcagag 1208
          || ||  |||| ||| | |||||   |||||||| |   || ||  | || |||   |||
Db    122 CGCAAGGCCATCGACTACATTCGCTTTCTGCAACACAGCAACCAGAAACTCAAGCAGGAG 181

Qy   1209 ttgcagcgcctgc 1221
             |   | ||||
Db    182 AACCTAAGTCTGC 194


RESULT  10
AW701427
LOCUS       AW701427      506 bp    mRNA            EST       18-APR-2000
DEFINITION  uq86a04.y1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:2937966 5'
            similar to gb:U00968 STEROL REGULATORY ELEMENT BINDING PROTEIN 1
            (HUMAN);, mRNA sequence.
ACCESSION   AW701427
VERSION     AW701427.1  GI:7585558
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 506)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 6, 2000 this sequence version replaced gi:6677089.
            Other_ESTs: uq86a04.x1
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml

            MGI:1050386
            Seq primer: -40RP from Gibco
            High quality sequence stop: 449.
FEATURES             Location/Qualifiers
     source          1. .506
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2937966"
                     /clone_lib="NCI_CGAP_Lu33"
                     /tissue_type="pooled lung tumors"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st
                     strand cDNA was prepared from mRNA obtained from pooled
                     lung tumors with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCTGTTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      130 a    146 c    142 g     88 t
ORIGIN


  Query Match             2.4%;  Score 91.8;  DB 80;  Length 506;
  Best Local Similarity   66.3%;  Pred. No. 5.3e-14;
  Matches  132;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy   1029 aagcgctcggcccacaacgccatcgagcggcgctatcgcacctcaatcaacgacaagatt 1088
          |||||| | |||||||| ||||| |||  |||||| ||  |||| ||||| |||||||||
Db     52 AAGCGCACAGCCCACAATGCCATTGAGAAGCGCTACCGGTCCTCTATCAATGACAAGATT 111

Qy   1089 aacgagttgaagaacttggtagtgggagagcaggccaagctgaacaagtccgcagtgttg 1148
             ||| | ||  || |||| |||||     | || ||||||||| || || || || |||
Db    112 GTGGAGCTCAAAGACCTGGTGGTGGGCACTGAAGCAAAGCTGAATAAATCTGCTGTCTTG 171

Qy   1149 cggaaatccatagacaagattcgggatctgcaacgccagaatcacgatctgaaggcagag 1208
          || ||  |||| ||| | || ||     |||| | |   || ||  | || |||   |||
Db    172 CGCAAGGCCATCGACTACATCCGCTTCTTGCAGCACAGCAACCAGAAGCTCAAGCAGGAG 231

Qy   1209 ttgcagcgcctgcagaggg 1227
             | |  ||| |  || |
Db    232 AACCTGACCCTACGAAGTG 250
```

```
RESULT  11
AJ396343
LOCUS       AJ396343      574 bp    mRNA            EST       01-MAR-2000
DEFINITION  AJ396343 dkfz426 Gallus gallus cDNA clone 26c11r1, mRNA sequence.
ACCESSION   AJ396343
VERSION     AJ396343.1  GI:7128184
KEYWORDS    EST.
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 574)
  AUTHORS   Buerstedde,J.M.
  TITLE     A large database of chicken bursal ESTs as a resource of the
            analysis of vertebrate gene function
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Buerstedde JM
            Cellular Immunology
            Heinrich-Pette-Institute
            Martinistr. 52, 20251 Hamburg, Germany
            Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
FEATURES             Location/Qualifiers
     source          1. .574
                     /organism="Gallus gallus"
                     /strain="CB"
                     /db_xref="taxon:9031"
                     /clone="26c11r1"
                     /clone_lib="dkfz426"
                     /tissue_type="Bursa of Fabricius"
BASE COUNT      127 a    196 c    170 g     81 t
ORIGIN
```

Query Match 2.4%; Score 89.6; DB 47; Length 574;
Best Local Similarity 51.7%; Pred. No. 2.2e-13;
Matches 243; Conservative 0; Mismatches 209; Indels 18; Gaps 1;

```
Qy   3041 ggaactggaacacatgaatatggaggacgatggcttctaccaagtgccaggtgaagtgct 3100
          | | | ||    ||  |    || |   | ||    |||||| |   ||       | |||
Db     16 GAAGCAGGCGAGCAGCAGCCAGGCGCTGGGTGAGACCTACCATGCCTCACCCCCCGAGCT 75

Qy   3101 cgagaagttccagaccgatttgaactcgttgcgcaacattgtggagaatataccgaacgc 3160
          |      ||||||   ||  |   |    |||||||  | |   |      |   |  |||
Db     76 CACTGGCTTCCAGCGTGACCTCGGCAGCCTGCGCAAGCTGGCCCACGGCTTCAGGCCCGC 135

Qy   3161 ccaatcgcgcatatatttgtacgaggcagtttgtcgcctgatggctggagcctcaccgtg 3220
            |         | |   |  |||||||       ||||||||||| || |||   ||
Db    136 GTACCGCAAGGTCTTCCTTCACGAGGCCACCGTGCGCCTGATGGCGGGGGCCAGCCCTAC 195

Qy   3221 tccaacgcaacagctcttggacaggagtctgcgatcacgcaacgcccactcgtccatctt 3280
            |  || || |||||  ||||    |||||  |    ||||   ||||
Db    196 CCGCACCCACCAGCTGCTGGAGCACAGTCTCAGGCGTCGCACAACCCAG---------- 244

Qy   3281 ctgcggcagcaaggatcggcggcagcagaacttcgtgggcggagagcgggaacgggcttc 3340
               | ||   | |   || ||| | ||    ||   ||  || |||| || ||  |
Db    245 -------AACAGCAAGCAAGGGGAGCTGGACACGCTGCCAGGGCAGAGGGAGCGAGCCAC 297

Qy   3341 ggccatgtacgtggcctgcaagtatctcccgcctgcgctgctcagctccccgggtgaacg 3400
           |||||     ||||||||    | |||||  |   |  | ||   ||| ||  |  | ||
Db    298 TGCCATCCTGCTGGCCTGCCGCCACCTCCCACTCTCTTTCCTGTCCTCGCCCAGCCAGCG 357

Qy   3401 tgctggcatgttagccgaggcggccaagaccctggagaaggtgggcgacaagcgaaagct 3460
          ||| | | || | || ||||| ||    || ||||||||||||||| |||||||   |
Db    358 TGCCGTCCTGCTGGCTGAGGCTGCACGCACGCTGGAGAAGGTGGGGGACAAGCGCTCGTG 417

Qy   3461 caaggagtgctaccagctgatgaagtcgctgggcaacggcattggcagcg 3510
          ||| || ||| | ||| ||||      |||| |    ||||  | || ||
Db    418 CAACGACTGCCAACAGATGATTGTCAAGCTGAGTGGGGGCACAGCCATCG 467
```

```
RESULT  12
AL040138
LOCUS       AL040138      554 bp    mRNA            EST       29-FEB-2000
DEFINITION  DKFZp434D1913_r1 434 (synonym: htes3) Homo sapiens cDNA clone
            DKFZp434D1913 5', mRNA sequence.
ACCESSION   AL040138
VERSION     AL040138.1  GI:5409105
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 554)
  AUTHORS   Koehrer,K., Beyer,A., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
  TITLE     EST (Koehrer, et al.)
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 15, 1998 this sequence version replaced gi:3224028.
            Contact: Koehrer K
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by BMFZ (Biomedical Research Center at the Charite,
            Berlin/Germany) within the cDNA sequencing consortium of the German
            Genome Project.
            No s1 sequence available.
            This clone (DKFZp434D1913) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .554
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp434D1913"
                     /clone_lib="434 (synonym: htes3)"
                     /tissue_type="testis"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT      155 a    153 c    138 g    108 t
ORIGIN
```

Query Match 2.1%; Score 80.6; DB 47; Length 554;
Best Local Similarity 59.3%; Pred. No. 6e-11;
Matches 137; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

```
Qy    997 tcaaccgggttcaacccaaggtgaaggaagtaaagcgctcggcccacaacgccatcgagc 1056
          |||| | | || | |||       | | || || ||| | |||| ||   ||| |||
Db    170 TCAAGCAGCTTGAGCCCCCCAAAGAAGGAGAAAGGCGGACAACCCATAATATCATTGAGA 229

Qy   1057 ggcgctatcgcacctcaatcaacgacaagattaacgagttgaagaacttggtagtgggag 1116
           || |||||| |||| ||||| ||||| || | ||| |||||  || ||||  ||||
Db    230 AACGATATCGCTCCTCCATCAATGACAAAATCATCGAATTGAAAGACCTGGTCATGGGGA 289

Qy   1117 agcaggccaagctgaacaagtccgcagtgttgcggaaatccatagacaagattcgggatc 1176
             | |||||| || ||||||| |  ||  || ||||  |||| ||  | ||       |
Db    290 CAGACGCCAAGATGCACAAGTCTGGCGTTCTGAGGAAGGCCATTGATTACATCAAATACT 349

Qy   1177 tgcaacgccagaatcacgatctgaaggcagagttgcagcgcctgcagaggg 1227
          |||| |     |||||  | |||      |||      |   ||| ||  ||
Db    350 TGCAGCAGGTCAATCATAAACTGCGCCAGGAGAACATGGTGCTGAAGCTGG 400
```

```
RESULT  13
AA806181/c
LOCUS       AA806181      443 bp    mRNA            EST       25-MAR-1998
DEFINITION  oc25e08.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1350758 3'
            similar to gb:U00968 STEROL REGULATORY ELEMENT BINDING PROTEIN 1
            (HUMAN);, mRNA sequence.
ACCESSION   AA806181
VERSION     AA806181.1  GI:2874931
```

```
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 443)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 19, 1997 this sequence version replaced gi:1520533.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 784    Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 420.
FEATURES             Location/Qualifiers
     source          1. .443
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1350758"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT       72 a    125 c     96 g    150 t
ORIGIN


  Query Match             1.9%;  Score 73;  DB 31;  Length 443;
  Best Local Similarity   63.3%;  Pred. No. 6.6e-09;
  Matches  112;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy   1021 aggaagtaaagcgctcggcccacaacgccatcgagcggcgctatcgcacctcaatcaacg 1080
           | | || || ||| |  |||| ||   ||| |||   || |||||| |||| ||||| |
Db    433 AAGGAGAAAGGCGGACAACCCATAATATCATTGAGAAACGATATCGCTCCTCCATCAATG 374

Qy   1081 acaagattaacgagttgaagaacttggtagtgggagagcaggccaagctgaacaagtccg 1140
           |||| || | ||| |||||  || ||||  ||||     | |||||| || |||||||
Db    373 ACAAAATCATCGAATTGAAAGACCTGGTCATGGGGACAGACGCCAAGATGCACAAGTCTG 314

Qy   1141 cagtgttgcggaaatccatagacaagattcgggatctgcaacgccagaatcacgatc 1197
             ||  || ||||  |||| ||  | |    |    ||| || |||   ||  |  ||
Db    313 GCGTTCTGAGGAAGGCCATTGATTACACCAGATGCCTGAAAGGCCTCCATGGCTTTC 257


RESULT  14
AI386259
LOCUS       AI386259      524 bp    mRNA            EST       27-JAN-1999
DEFINITION  mm45e05.y1 Stratagene mouse melanoma (#937312) Mus musculus cDNA
            clone IMAGE:524480 5' similar to gb:U00968 STEROL REGULATORY
            ELEMENT BINDING PROTEIN 1 (HUMAN);, mRNA sequence.
ACCESSION   AI386259
VERSION     AI386259.1  GI:4199722
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 524)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:318328
            This read is a RESEQUENCE of a previously sequenced mouse clone
            This read has been verified (found to hit its original self in the
            correct orientation)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 406.
FEATURES             Location/Qualifiers
     source          1. .524
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:524480"
                     /clone_lib="Stratagene mouse melanoma (#937312)"
                     /tissue_type="melanoma"
                     /dev_stage="M2 cells"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. From M2 cells, a highly metastatic derivative of
                     the K-1735 (mouse) melanoma. Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
BASE COUNT       96 a    168 c    165 g     94 t      1 others
ORIGIN


  Query Match             1.9%;  Score 70.8;  DB 39;  Length 524;
  Best Local Similarity   50.0%;  Pred. No. 2.7e-08;
  Matches  211;  Conservative    0;  Mismatches  202;  Indels    9;  Gaps    1;

Qy   3080 ccaagtgccaggtgaagtgctcgagaagttccagaccgatttgaactcgttgcgcaacat 3139
           ||| |   | | |   | |||       |||||    ||  ||| |    ||||     |
Db     87 CCAGGCCTCTGCTCTGGAGCTGCGTGGTTTCCAACATGACCTGAGCAGCCTGCGGCGGTT 146

Qy   3140 tgtggagaatataccgaacgcccaatcgcgcatatatttgtacgaggcagtttgtcgcct 3199
             |   |||   | | |   ||      | |  |||   |  | |||||      ||| ||
Db    147 GGCACAGAGCTTCCGGCCTGCTATGAGGAGGGTATTCCTACATGAGGCCACAGCTCGGCT 206

Qy   3200 gatggctggagcctcaccgtgtccaacgcaacagctcttggacaggagtctgcgatcacg 3259
           |||||| |||||    ||    |  || || |||||| ||||  | ||||||  |    |
Db    207 GATGGCAGGAGCAAGTCCTGCCCGGACACACCAGCTCCTGGATCGCAGTCTGAGGAGGAG 266

Qy   3260 caacgcccactcgtccatcttctgcggcagcaaggatcggcggcagcagaacttcgtggg 3319
               |    |     ||      ||         || | |  ||  |          |
Db    267 GGCAGGTTCCAGTGGCAAAGGAGGCACTGCAGCTGAGCTGGAGCCAC---------GGCC 317

Qy   3320 cggagagcgggaacgggcttcggccatgtacgtggcctgcaagtatctcccgcctgcgct 3379
          |  |  |||||| |   |   |||| ||    |||| | |   |||| || ||||  |
Db    318 CACATGGCGGGAGCACACCGAGGCCCTGCTGTTGGCATCCTGCTATCTGCCCCCTGCCTT 377
```

```
Qy   3380 gctcagctccccgggtgaacgtgctggcatgttagccgaggcggccaagaccctggagaa 3439
          ||      | || ||  | ||      ||||| | ||||||||||    ||| | |||||
Db    378 CCTGTCGGCTCCTGGGCAGCGAATGAGCATGCTGGCCGAGGCGGCACGCACCGTAGAGAA 437

Qy   3440 ggtgggcgacaagcgaaagctcaaggagtgctaccagctgatgaagtcgctgggcaacgg 3499
          | | |||||  | ||    ||   ||| ||| | ||| || |   |   ||||||  |||
Db    438 GCTTGGCGATCACCGGCTACTGCTGGACTGCCAGCAGATGCTCCTGCGCCTGGGCGGCGG 497

Qy   3500 ca 3501
          |
Db    498 AA 499


RESULT  15
AA823111
LOCUS       AA823111      412 bp    mRNA            EST       17-FEB-1998
DEFINITION  vw40g09.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:1246336 5' similar to gb:U00968 STEROL REGULATORY ELEMENT
            BINDING PROTEIN 1 (HUMAN);, mRNA sequence.
ACCESSION   AA823111
VERSION     AA823111.1  GI:2892979
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 412)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Jan 5, 1998 this sequence version replaced gi:2581046.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:660024
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 367.
FEATURES             Location/Qualifiers
     source          1. .412
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1246336"
                     /clone_lib="Soares_mammary_gland_NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT       69 a    146 c    114 g     83 t
ORIGIN
```

```
  Query Match             1.9%;  Score 70.4;  DB 31;  Length 412;
  Best Local Similarity   61.4%;  Pred. No. 3.3e-08;
  Matches  113;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy   3321 ggagagcgggaacgggcttcggccatgtacgtggcctgcaagtatctcccgcctgcgctg 3380
          ||  | || || || ||| | |||||     ||||||||    |||| || |   |  |
Db    170 GGCCAACGAGAGCGAGCTACTGCCATCCTGCTGGCCTGCCGCCATCTTCCCCTCTCTTTC 229

Qy   3381 ctcagctccccgggtgaacgtgctggcatgttagccgaggcggccaagaccctggagaaag 3440
          ||    |||||||  | || || |   || | || || || |||   ||||||||||||
Db    230 CTGTCTTCCCCGGGCCAGCGGGCAGTGCTGCTGGCTGAAGCTGCCCGCACCCTGGAGAAG 289

Qy   3441 gtgggcgacaagcgaaagctcaaggagtgctaccagctgatgaagtcgctgggcaacggc 3500
          ||||| |||   ||      ||  || ||| | ||| ||||      |||||||   |||
Db    290 GTGGGTGACCGCCGTTCCTGCAGTGACTGCCAGCAGATGATTGTCAAGCTGGGCGGTGGC 349

Qy   3501 attg 3504
          | ||
Db    350 ACTG 353


Search completed: September  4, 2000, 00:25:40
Job time: 19915 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2000, 00:00:24 ; Search time 11551.9 Seconds
(without alignments)
637.752 Million cell updates/sec

Title: US-09-332-522B-5
Perfect score: 4128
Sequence: 1 gtgtgcctgactgttttgta.....tgactaaatgttaaaataaa 4128

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: em_fun:*
17: em_hum1:*
18: em_hum2:*
19: em_in:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_sy:*
29: em_un:*
30: em_vi:*
31: gb_htg1:*
32: gb_htg2:*
33: gb_in1:*
34: gb_in2:*
35: em_ba1:*
36: em_ba2:*
37: em_hum3:*
38: em_hum4:*
39: gb_pr4:*
40: gb_htg3:*
41: gb_htg4:*
42: gb_htg5:*
43: gb_htg6:*
44: gb_htg7:*
45: em_htg1:*
46: em_htg2:*
47: em_htg3:*
48: em_hum5:*
49: gb_pl3:*
50: gb_pr5:*
51: gb_htg8:*
52: gb_htg9:*
53: gb_htg10:*
54: gb_htg11:*
55: gb_htg12:*
56: gb_htg13:*
57: gb_htg14:*
58: gb_in3:*
59: gb_htg15:*
60: gb_htg16:*
61: gb_htg17:*
62: em_htg4:*
63: em_htg5:*
64: em_htg6:*
65: em_htg7:*
66: em_hum6:*
67: gb_htg18:*
68: gb_htg19:*
69: gb_htg20:*
70: gb_htg21:*
71: gb_htg22:*
72: gb_htg23:*
73: gb_htg24:*
74: gb_htg25:*
75: gb_htg26:*
76: gb_htg27:*
77: gb_htg28:*
78: gb_htg29:*
79: gb_htg30:*
80: gb_htg31:*
81: gb_vi1:*
82: gb_vi2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	3279.6	79.4	67188	45	AC005424	Ac005424 *** SEQUE
	2	3279.6	79.4	135182	34	AC007121	AC007121 Drosophil
	3	3279.6	79.4	313072	58	AE003784	AE003784 Drosophil
	4	2390.4	57.9	60625	43	AC020090	AC020090 Drosophil
	5	1126.4	27.3	12978	42	AC014229	AC014229 Drosophil
	6	1017.4	24.6	109699	42	AC009255	AC009255 Drosophil
	7	263.2	6.4	790	13	CNS01P6Y	AL154442 Anopheles
	8	193.6	4.7	3976	9	D83782	D83782 Human mRNA
	9	187.4	4.5	4119	12	CGU67060	U67060 Cricetulus
	10	153.2	3.7	118474	41	AC008339	AC008339 Drosophil
c	11	137	3.3	109699	42	AC009255	AC009255 Drosophil
c	12	62.4	1.5	10844	42	AC017232	AC017232 Drosophil
c	13	62.4	1.5	28380	34	AE003066	AE003066 Drosophil
	14	61.4	1.5	16047	42	AC017996	AC017996 Drosophil
	15	61.4	1.5	311786	58	AE003781	AE003781 Drosophil
c	16	61.2	1.5	95668	67	AC009254	AC009254 Drosophil
	17	60.4	1.5	40940	33	CED2013	Z47808 Caenorhabdi
	18	60.4	1.5	191908	31	CEY51B9	Z98947 Caenorhabdi
	19	59.6	1.4	40400	42	AC012915	AC012915 Drosophil
c	20	59.6	1.4	158322	51	AC011759	AC011759 Drosophil
c	21	59.6	1.4	274456	34	AE003574	AE003574 Drosophil
	22	58.8	1.4	8132	42	AC018078	AC018078 Drosophil
	23	58.8	1.4	33240	34	AE003021	AE003021 Drosophil
	24	57.8	1.4	33013	34	DMU87286	U87286 Drosophila

c	25	56.4	1.4	8516	42	AC018056	AC018056 Drosophil
	26	56.4	1.4	31062	34	AE002734	AE002734 Drosophil
c	27	56.4	1.4	199621	54	AC023230	AC023230 Homo sapi
	28	56.2	1.4	5991	42	AC014934	AC014934 Drosophil
c	29	56.2	1.4	316355	58	AE003843	AE003843 Drosophil
	30	55.6	1.3	3078	42	AC017299	AC017299 Drosophil
	31	55.6	1.3	20775	34	AE002872	AE002872 Drosophil
c	32	54.6	1.3	5235	42	AC015010	AC015010 Drosophil
c	33	54.6	1.3	114299	58	AE003845	AE003845 Drosophil
c	34	54.2	1.3	3505	42	AC018141	AC018141 Drosophil
c	35	54.2	1.3	311786	58	AE003781	AE003781 Drosophil
c	36	53.6	1.3	6207	42	AC017165	AC017165 Drosophil
	37	53.6	1.3	67870	34	AE002910	AE002910 Drosophil
c	38	53	1.3	35184	42	AC014624	AC014624 Drosophil
c	39	53	1.3	105037	58	AE003847	AE003847 Drosophil
	40	52.4	1.3	6609	42	AC018334	AC018334 Drosophil
c	41	52.4	1.3	30869	34	AE002759	AE002759 Drosophil
c	42	52	1.3	7932	42	AC014648	AC014648 Drosophil
	43	52	1.3	15744	42	AC014618	AC014618 Drosophil
c	44	52	1.3	41037	42	AC017476	AC017476 Drosophil
	45	52	1.3	48870	53	AC010838	AC010838 Drosophil

ALIGNMENTS

```
RESULT   1
AC005424
ID   AC005424   standard; DNA; HTG; 67188 BP.
XX
AC   AC005424;
XX
SV   AC005424.1
XX
DT   17-AUG-1998 (Rel. 56, Created)
DT   18-DEC-1998 (Rel. 58, Last updated, Version 7)
XX
DE   *** SEQUENCING IN PROGRESS *** Drosophila melanogaster; Chromosome
DE   2R; Region 42A8-42A16; P1 clone DS06954; HTGS phase 1, 6 unordered
DE   pieces.
XX
KW   HTG; HTGS_PHASE1.
XX
OS   Drosophila melanogaster (fruit fly)
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
OC   Drosophilidae; Drosophila.
XX
RN   [1]
RP   1-67188
RA   Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
RA   Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA   Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA   Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
RA   Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA   Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Punch E.,
RA   Sequeira A., Sethi H., Snir E., Svirskas R.R., Twomey B., Wan K.H.,
RA   Weinburg T., Zhang R., Zieran L.L., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster";
RL   Unpublished.
XX
RN   [2]
RP   1-67188
RA   Celniker S.E., George R.A., Galle R., Svirskas R.R., Hoskins R.A.,
RA   Agbayani A., Arcaina T.T., Baxter E., Blazej R.G., Chavez C., Chew M.,
RA   Doyle C.M., Farfan D.E., Flanagan J., Houston K.A., Hummasti S.R.,
RA   Karra K., Kearney L., Kim S.H., Lee B., Lomotan M.A., Mak J., Mazda P.,
RA   Mok M.S., Moshrefi A.R., Moshrefi M., Nixon K., Pacleb J.M., Park S.,
RA   Pfeiffer B., Punch E., Snir E., Twomey B., Wan K.H., Whitelaw K.R., Yee A.,
RA   Zhang R., Zieran L.L., Kimmel B.E.;
RT   ;
RL   Submitted (14-AUG-1998) to the EMBL/GenBank/DDBJ databases.
RL   Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121,
RL   Berkeley, CA 94720, USA
XX
CC   On Dec 17, 1998 this sequence version replaced gi:3893038.
CC   For further information about this sequence, including its location
CC   and relationship to other sequences, please visit our sequence
CC   archive Web site (http://www.fruitfly.org/sequence/) or send email
CC   to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
CC   the following cutoffs: length >= 400 bases, phrap computed error
CC   rate <= 1/10. P1 library location: 73-42.
CC   ***                                                          ***
CC   *** WARNING: Phase 1 High Throughput Genome Sequence ***
CC   ***                                                          ***
CC   * This sequence is unfinished. It consists of 6 contigs for
CC   * which the order is not known; their order in this record is
CC   * arbitrary. In some cases, the exact lengths of the gaps
CC   * between the contigs are also unknown; these gaps are  presented
CC   * as runs of N as a convenience only. When sequencing is complete,
CC   * the sequence data presented in this record will be replaced
CC   *by a single finished sequence with the same accession number.
CC   *        1     1168: contig of 1168 bp in length
CC   *     1169     1248: gap of unknown length
CC   *     1249     2019: contig of 771 bp in length
CC   *     2020     2099: gap of unknown length
CC   *     2100     3241: contig of 1142 bp in length
CC   *     3242     3321: gap of unknown length
CC   *     3322     5033: contig of 1712 bp in length
CC   *     5034     5113: gap of unknown length
CC   *     5114     7443: contig of 2330 bp in length
CC   *     7444     7523: gap of unknown length
CC   *     7524    67188: contig of 59665 bp in length.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .67188
FT                   /chromosome="2"
FT                   /db_xref="taxon:7227"
FT                   /organism="Drosophila melanogaster"
FT                   /strain="y2; cn bw sp"
FT                   /clone="DS06954 (D338)"
FT                   /clone_lib="P1 library, partial Sau3A in pAd10sacBII"
FT                   /map="42A8-42A16"
XX
SQ   Sequence 67188 BP; 19284 A; 14167 C; 14128 G; 19185 T; 424 other;


  Query Match             79.4%;  Score 3279.6;  DB 45;  Length 67188;
  Best Local Similarity   96.4%;  Pred. No. 0;
  Matches 3423;  Conservative    0;  Mismatches    9;  Indels  118;  Gaps    2;

Qy     697 acgatatttcagtaccataacctacagaaatcaaaagtttccgcggcggaaatgctgttt 756
           | | |  | || |  || ||||||||||||||||||||||||||||||||||||||||||
Db   63447 ATGTTTGTCCATTTTCAGAACCTACAGAAATCAAAAGTTTCCGCGGCGGAAATGCTGTTT 63506

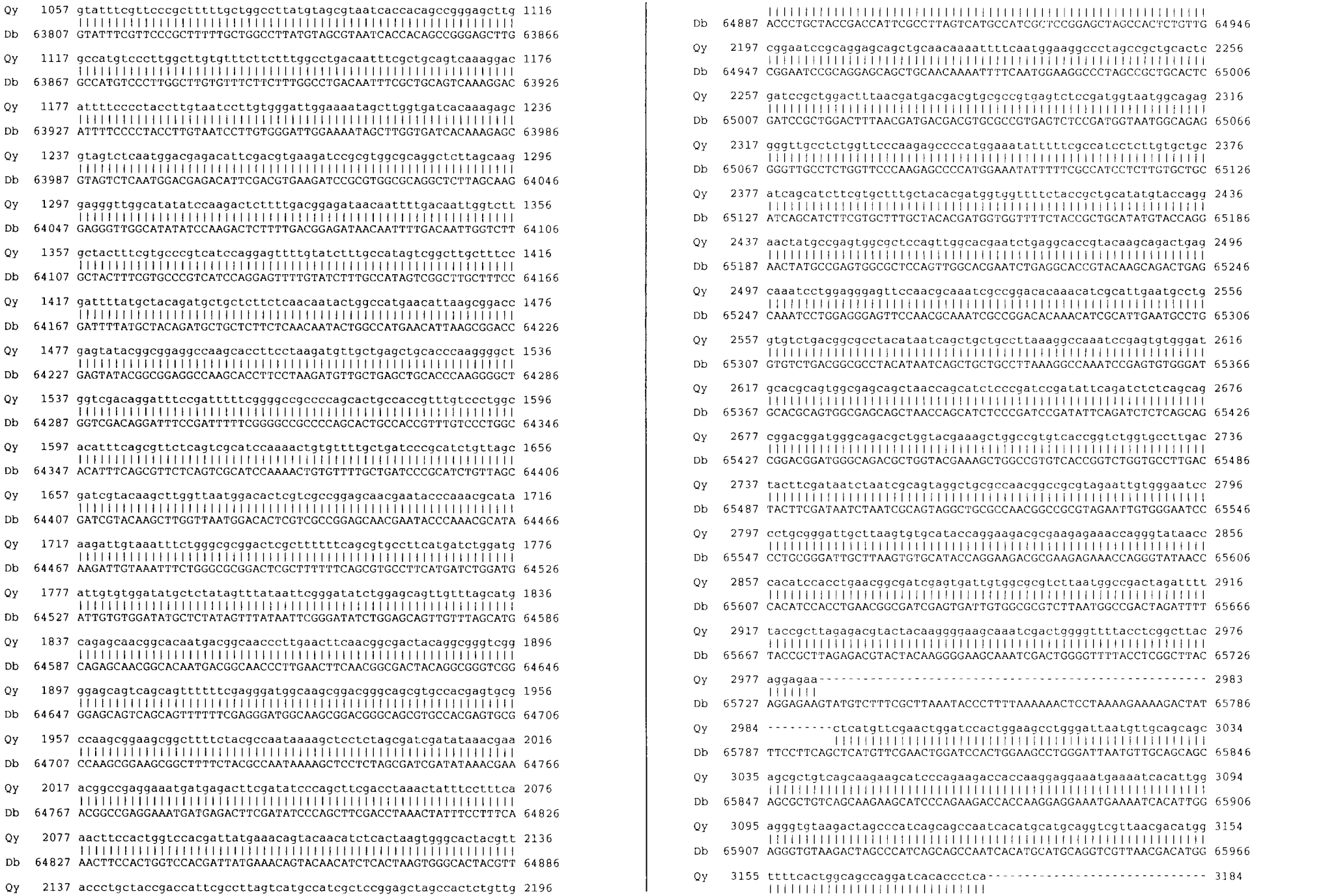
Qy     757 ggattacccatgcaggacactggattcaagcgctatccattgcgcgctcggtcgcgtatt 816
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63507 GGATTACCCATGCAGGACACTGGATTCAAGCGCTATCCATTGCGCGCTCGGTCGCGTATT 63566

Qy     817 atacagtatgccttgacgttattcctcaagcacaacgatatggagtatctggacactcta 876
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63567 ATACAGTATGCCTTGACGTTATTCCTCAAGCACAACGATATGGAGTATCTGGACACTCTA 63626

Qy     877 aaggaaaagctgctgcgacactatccccccactcccgttggctagtgcgtcggctgaagag 936
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63627 AAGGAAAAGCTGCTGCGACACTATCCCCCACTCCCGTTGGCTAGTGCGTCGGCTGAAGAG 63686

Qy     937 ccgacgaccataacttacatcttttatccaggagagtacaggatgtgggagctggtgcct 996
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63687 CCGACGACCATAACTTACATCTTTTATCCAGGAGAGTACAGGATGTGGGAGCTGGTGCCT 63746

Qy     997 tacacagtggcctttatgttggtgtttgcttatgtgtacttctctgttcgaaaaatcgat 1056
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63747 TACACAGTGGCCTTTATGTTGGTGTTTGCTTATGTGTACTTCTCTGTTCGAAAAATCGAT 63806
```


Qy 1057 gtatttcgttcccgctttttgctggccttatgtagcgtaatcaccacagccgggagcttg 1116
Db 63807 GTATTTCGTTCCCGCTTTTTGCTGGCCTTATGTAGCGTAATCACCACAGCCGGGAGCTTG 63866
Qy 1117 gccatgtcccttggcttgtgtttcttctttggcctgacaatttcgctgcagtcaaaggac 1176
Db 63867 GCCATGTCCCTTGGCTTGTGTTTCTTCTTTGGCCTGACAATTTCGCTGCAGTCAAAGGAC 63926
Qy 1177 attttcccctaccttgtaatccttgtgggattggaaaatagcttggtgatcacaaagagc 1236
Db 63927 ATTTTCCCCTACCTTGTAATCCTTGTGGGATTGGAAAATAGCTTGGTGATCACAAAGAGC 63986
Qy 1237 gtagtctcaatggacgagacattcgacgtgaagatccgcgtggcgcaggctcttagcaag 1296
Db 63987 GTAGTCTCAATGGACGAGACATTCGACGTGAAGATCCGCGTGGCGCAGGCTCTTAGCAAG 64046
Qy 1297 gaggggttggcatatatccaagactcttttgacggagataacaattttgacaattggtctt 1356
Db 64047 GAGGGTTGGCATATATCCAAGACTCTTTTGACGGAGATAACAATTTTGACAATTGGTCTT 64106
Qy 1357 gctactttcgtgcccgtcatccaggagttttgtatctttgccatagtcggcttgctttcc 1416
Db 64107 GCTACTTTCGTGCCCGTCATCCAGGAGTTTTGTATCTTTGCCATAGTCGGCTTGCTTTCC 64166
Qy 1417 gattttatgctacagatgctgctcttctcaacaatactggccatgaacattaagcggacc 1476
Db 64167 GATTTTATGCTACAGATGCTGCTCTTCTCAACAATACTGGCCATGAACATTAAGCGGACC 64226
Qy 1477 gagtatacggcggaggccaagcaccttcctaagatgttgctgagctgcacccaaggggct 1536
Db 64227 GAGTATACGGCGGAGGCCAAGCACCTTCCTAAGATGTTGCTGAGCTGCACCCAAGGGGCT 64286
Qy 1537 ggtcgacaggatttccgatttttcggggccgccccagcactgccaccgtttgtccctggc 1596
Db 64287 GGTCGACAGGATTTCCGATTTTTCGGGGCCGCCCCAGCACTGCCACCGTTTGTCCCTGGC 64346
Qy 1597 acatttcagcgttctcagtcgcatccaaaactgtgttttgctgatcccgcatctgttagc 1656
Db 64347 ACATTTCAGCGTTCTCAGTCGCATCCAAAACTGTGTTTTGCTGATCCCGCATCTGTTAGC 64406
Qy 1657 gatcgtacaagcttggttaatggacactcgtcgccggagcaacgaatacccaaacgcata 1716
Db 64407 GATCGTACAAGCTTGGTTAATGGACACTCGTCGCCGGAGCAACGAATACCCAAACGCATA 64466
Qy 1717 aagattgtaaatttctgggcgcggactcgctttttttcagcgtgccttcatgatctggatg 1776
Db 64467 AAGATTGTAAATTTCTGGGCGCGGACTCGCTTTTTTCAGCGTGCCTTCATGATCTGGATG 64526
Qy 1777 attgtgtggatatgctctatagtttataattcgggatatctggagcagttgtttagcatg 1836
Db 64527 ATTGTGTGGATATGCTCTATAGTTTATAATTCGGGATATCTGGAGCAGTTGTTTAGCATG 64586
Qy 1837 cagagcaacggcacaatgacggcaacccttgaacttcaacggcgactacaggcgggtcgg 1896
Db 64587 CAGAGCAACGGCACAATGACGGCAACCCTTGAACTTCAACGGCGACTACAGGCGGGTCGG 64646
Qy 1897 ggagcagtcagcagttttttcgagggatggcaagcggacgggcagcgtgccacgagtgcg 1956
Db 64647 GGAGCAGTCAGCAGTTTTTTCGAGGGATGGCAAGCGGACGGGCAGCGTGCCACGAGTGCG 64706
Qy 1957 ccaagcggaagcggcttttctacgccaataaaagctcctctagcgatcgatataaacgaa 2016
Db 64707 CCAAGCGGAAGCGGCTTTTCTACGCCAATAAAAGCTCCTCTAGCGATCGATATAAACGAA 64766
Qy 2017 acggccgaggaaatgatgagacttcgatatcccagcttcgacctaaactatttccttca 2076
Db 64767 ACGGCCGAGGAAATGATGAGACTTCGATATCCCAGCTTCGACCTAAACTATTTCCTTTCA 64826
Qy 2077 aacttccactggtccacgattatgaaacagtacaacatctcactaagtgggcactacgtt 2136
Db 64827 AACTTCCACTGGTCCACGATTATGAAACAGTACAACATCTCACTAAGTGGGCACTACGTT 64886
Qy 2137 accctgctaccgaccattcgccttagtcatgccatcgctccggagctagccactctgttg 2196
Db 64887 ACCCTGCTACCGACCATTCGCCTTAGTCATGCCATCGCTCCGGAGCTAGCCACTCTGTTG 64946
Qy 2197 cggaatccgcaggagcagctgcaacaaaatttttcaatggaaggccctagccgctgcactc 2256
Db 64947 CGGAATCCGCAGGAGCAGCTGCAACAAAATTTTCAATGGAAGGCCCTAGCCGCTGCACTC 65006
Qy 2257 gatccgctggactttaacgatgacgacgtgcgccgtgagtctccgatggtaatggcagag 2316
Db 65007 GATCCGCTGGACTTTAACGATGACGACGTGCGCCGTGAGTCTCCGATGGTAATGGCAGAG 65066
Qy 2317 gggttgcctctggttcccaagagccccatggaaatatttttcgccatcctcttgtgctgc 2376
Db 65067 GGGTTGCCTCTGGTTCCCAAGAGCCCCATGGAAATATTTTTCGCCATCCTCTTGTGCTGC 65126
Qy 2377 atcagcatcttcgtgctttgctacacgatggtggttttctaccgctgcatatgtaccagg 2436
Db 65127 ATCAGCATCTTCGTGCTTTGCTACACGATGGTGGTTTTCTACCGCTGCATATGTACCAGG 65186
Qy 2437 aactatgccgagtggcgctccagttggcacgaatctgaggcaccgtacaagcagactgag 2496
Db 65187 AACTATGCCGAGTGGCGCTCCAGTTGGCACGAATCTGAGGCACCGTACAAGCAGACTGAG 65246
Qy 2497 caaatcctggagggagttccaacgcaaatcgccggacacaaacatcgcattgaatgcctg 2556
Db 65247 CAAATCCTGGAGGGAGTTCCAACGCAAATCGCCGGACACAAACATCGCATTGAATGCCTG 65306
Qy 2557 gtgtctgacggcgcctacataatcagctgctgccttaaaggccaaatccgagtgtgggat 2616
Db 65307 GTGTCTGACGGCGCCTACATAATCAGCTGCTGCCTTAAAGGCCAAATCCGAGTGTGGGAT 65366
Qy 2617 gcacgcagtggcgagcagctaaccagcatctcccgatccgatattcagatctctcagcag 2676
Db 65367 GCACGCAGTGGCGAGCAGCTAACCAGCATCTCCCGATCCGATATTCAGATCTCTCAGCAG 65426
Qy 2677 cggacggatgggcagacgctggtacgaaagctggccgtgtcaccggtctggtgccttgac 2736
Db 65427 CGGACGGATGGGCAGACGCTGGTACGAAAGCTGGCCGTGTCACCGGTCTGGTGCCTTGAC 65486
Qy 2737 tacttcgataatctaatcgcagtaggctgcgccaacggccgcgtagaattgtgggaatcc 2796
Db 65487 TACTTCGATAATCTAATCGCAGTAGGCTGCGCCAACGGCCGCGTAGAATTGTGGGAATCC 65546
Qy 2797 cctgcgggattgcttaagtgtgcataccaggaagacgcgaagagaaaccagggtataacc 2856
Db 65547 CCTGCGGGATTGCTTAAGTGTGCATACCAGGAAGACGCGAAGAGAAACCAGGGTATAACC 65606
Qy 2857 cacatccacctgaacggcgatcgagtgattgtggcgcgtcttaatggccgactagatttt 2916
Db 65607 CACATCCACCTGAACGGCGATCGAGTGATTGTGGCGCGTCTTAATGGCCGACTAGATTTT 65666
Qy 2917 taccgcttagagacgtactacaaggggaagcaaatcgactggggttttacctcggcttac 2976
Db 65667 TACCGCTTAGAGACGTACTACAAGGGGAAGCAAATCGACTGGGGTTTTACCTCGGCTTAC 65726
Qy 2977 aggagaa---------- 2983
Db 65727 AGGAGAAGTATGTCTTTCGCTTAAATACCCTTTTAAAAAACTCCTAAAAGAAAAGACTAT 65786
Qy 2984 ----------ctcatgttcgaactggatccactggaagcctgggattaatgttgcagcagc 3034
Db 65787 TTCCTTCAGCTCATGTTCGAACTGGATCCACTGGAAGCCTGGGATTAATGTTGCAGCAGC 65846
Qy 3035 agcgctgtcagcaagaagcatcccagaagaccaccaaggaggaaatgaaaatcacattgg 3094
Db 65847 AGCGCTGTCAGCAAGAAGCATCCCAGAAGACCACCAAGGAGGAAATGAAAATCACATTGG 65906
Qy 3095 agggtgtaagactagcccatcagcagccaatcacatgcatgcaggtcgttaacgacatgg 3154
Db 65907 AGGGTGTAAGACTAGCCCATCAGCAGCCAATCACATGCATGCAGGTCGTTAACGACATGG 65966
Qy 3155 ttttcactggcagccaggatcacaccctca---------- 3184

```
Db  65967 TTTTCACTGGCAGCCAGGATCACACCCTCAAGGTAGGCATGAAACATTTTATTAAATCAA 66026

Qy   3185 --------------------------aggtgtattgcctcaataagtcggatgttgagta 3218
                                    ||||||||||||||||||||||||||||||||||
Db  66027 CTTTCCTTACTCTAACTGGTCTCCATAGGTGTATTGCCTCAATAAGTCGGATGTTGAGTA 66086

Qy   3219 tacgctccacggtcactgtgggcctgtaacctgtctctttgtggatcgctggcaacctgg 3278
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66087 TACGCTCCACGGTCACTGTGGGCCTGTAACCTGTCTCTTTGTGGATCGCTGGCAACCTGG 66146

Qy   3279 cacaggggggtctgggtcccaggacggcctgctctgcgtatgggatctgttcacgggagc 3338
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66147 CACAGGGGGGTCTGGGTCCCAGGACGGCCTGCTCTGCGTATGGGATCTGTTCACGGGAGC 66206

Qy   3339 ctgcatgtataatatacaagctcacgacggagccgtcagctgcctggcctgtgcgcccag 3398
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66207 CTGCATGTATAATATACAAGCTCACGACGGAGCCGTCAGCTGCCTGGCCTGTGCGCCCAG 66266

Qy   3399 ttacgtaatctcgctaggcacggacgagaggatttgcgtatgggaacgatttcagggaaa 3458
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66267 TTACGTAATCTCGCTAGGCACGGACGAGAGGATTTGCGTATGGGAACGATTTCAGGGAAA 66326

Qy   3459 cctgttgactaccatcaacatctcaaacgcatactcgagcctactgatgctaacaccgtc 3518
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66327 CCTGTTGACTACCATCAACATCTCAAACGCATACTCGAGCCTACTGATGCTAACACCGTC 66386

Qy   3519 actattggttacgagcaaaatgggtaaggcctcattcttgattgccaatataagagggac 3578
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66387 ACTATTGGTTACGAGCAAAATGGGTAAGGCCTCATTCTTGATTGCCAATATAAGAGGGAC 66446

Qy   3579 agtaaataataaatttaattccaacacaggatctcttattgtgtgggatgtgcgcactgg 3638
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66447 AGTAAATAATAAATTTAATTCCAACACAGGATCTCTTATTGTGTGGGATGTGCGCACTGG 66506

Qy   3639 gcagccggctcgcgaggtcaaactggactttgcaaacctgcagctctgtcccaaaataat 3698
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66507 GCAGCCGGCTCGCGAGGTCAAACTGGACTTTGCAAACCTGCAGCTCTGTCCCAAAATAAT 66566

Qy   3699 gatgcttgcctgcgattcggtagtttgcgactacggaaatgagatccgcgtcgtccgctt 3758
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66567 GATGCTTGCCTGCGATTCGGTAGTTTGCGACTACGGAAATGAGATCCGCGTCGTCCGCTT 66626

Qy   3759 tcctatcgtggcagacaagtgccattaaagcgcaaaattttaatttagcgtggttcgcta 3818
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66627 TCCTATCGTGGCAGACAAGTGCCATTAAAGCGCAAAATTTTAATTTAGCGTGGTTCGCTA 66686

Qy   3819 gcacctaggaataagttgacttaaggctttaaaacgcctggaagtcattgacgcattcac 3878
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66687 GCACCTAGGAATAAGTTGACTTAAGGCTTTAAAACGCCTGGAAGTCATTGACGCATTCAC 66746

Qy   3879 tattttatataaatatatacactattagggtccgcagcaacttacggttttaacacaagc 3938
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66747 TATTTTATATAAATATATACACTATTAGGGTCCGCAGCAACTTACGGTTTTAACACAAGC 66806

Qy   3939 tgtacgtatctcatctctagaattttgtgttagtttgtggacactaagtgtaacagctac 3998
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66807 TGTACGTATCTCATCTCTAGAATTTTGTGTTAGTTTGTGGACACTAAGTGTAACAGCTAC 66866

Qy   3999 gctccggtaggttaaggaactaaactaaatgaatcagatatatacacatatattttcgcg 4058
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66867 GCTCCGGTAGGTTAAGGAACTAAACTAAATGAATCAGATATATACACATATATTTTCGCG 66926

Qy   4059 taattatataaactacatagtgtcttaaagcgcctcagcctaatataaaatgactaaatg 4118
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66927 TAATTATATAAACTACATAGTGTCTTAAAGCGCCTCAGCCTAATATAAAATGACTAAATG 66986

Qy   4119 ttaaaataaa 4128
          ||||||||||
Db  66987 TTAAAATAAA 66996


RESULT   2
AC007121
```

```
LOCUS       AC007121   135182 bp    DNA             INV       20-MAR-1999
DEFINITION  Drosophila melanogaster, chromosome 2R, region 42A8-42A16, P1
            clones DS06954 and DS05325, complete sequence.
ACCESSION   AC007121 AC005424 AC005426
VERSION     AC007121.1  GI:4454581
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 135182)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Sequencing of Drosophila chromosome 2R, region 42A8-42A16
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 135182)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-1999) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     On Mar 20, 1999 this sequence version replaced gi:4028926
            gi:4079609.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
            P1 library locations: 73-42, 56-45.
FEATURES             Location/Qualifiers
     source          1. .135182
                     /organism="Drosophila melanogaster"
                     /strain="y2; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
                     /map="42A8-42A16"
                     /clone_lib="P1 library, partial Sau3A in pAd10sacBII"
                     /clone="P1s DS06954 (D338) and DS05325 (D340)"
                     /note="DS06954 (D338) and DS05325 (D340) were submitted as
                     a project. DS05325 extends from P1 end at bp1 to P1 end at
                     bp 77,598. DS06954 extends from p1 end at bp 67,666 to p1
                     end at bp135,182.  "
BASE COUNT    39447 a  28166 c  28329 g  39240 t
ORIGIN


  Query Match             79.4%;  Score 3279.6;  DB 34;  Length 135182;
  Best Local Similarity   96.4%;  Pred. No. 0;
  Matches 3423;  Conservative    0;  Mismatches    9;  Indels  118;  Gaps    2;

Qy    697 acgatatttcagtaccataacctacagaaatcaaaagtttccgcggcggaaatgctgttt 756
          | | |  | || |  || ||||||||||||||||||||||||||||||||||||||||||
Db 131437 ATGTTTGTCCATTTTCAGAACCTACAGAAATCAAAAGTTTCCGCGGCGGAAATGCTGTTT 131496
```

```
Qy     757 ggattacccatgcaggacactggattcaagcgctatccattgcgcgctcggtcgcgtatt 816
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  131497 GGATTACCCATGCAGGACACTGGATTCAAGCGCTATCCATTGCGCGCTCGGTCGCGTATT 131556

Qy     817 atacagtatgccttgacgttattcctcaagcacaacgatatggagtatctggacactcta 876
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  131557 ATACAGTATGCCTTGACGTTATTCCTCAAGCACAACGATATGGAGTATCTGGACACTCTA 131616

Qy     877 aaggaaaagctgctgcgacactatcccccactcccgttggctagtgcgtcggctgaagag 936
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  131617 AAGGAAAAGCTGCTGCGACACTATCCCCCACTCCCGTTGGCTAGTGCGTCGGCTGAAGAG 131676

Qy     937 ccgacgaccataacttacatcttttatccaggagagtacaggatgtgggagctggtgcct 996
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  131677 CCGACGACCATAACTTACATCTTTTATCCAGGAGAGTACAGGATGTGGGAGCTGGTGCCT 131736

Qy     997 tacacagtggcctttatgttggtgtttgcttatgtgtacttctctgttcgaaaaatcgat 1056
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  131737 TACACAGTGGCCTTTATGTTGGTGTTTGCTTATGTGTACTTCTCTGTTCGAAAAATCGAT 131796

Qy    1057 gtatttcgttcccgctttttgctggccttatgtagcgtaatcaccacagccgggagcttg 1116
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  131797 GTATTTCGTTCCCGCTTTTTGCTGGCCTTATGTAGCGTAATCACCACAGCCGGGAGCTTG 131856

Qy    1117 gccatgtcccttggcttgtgtttcttctttggcctgacaatttcgctgcagtcaaaggac 1176
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  131857 GCCATGTCCCTTGGCTTGTGTTTCTTCTTTGGCCTGACAATTTCGCTGCAGTCAAAGGAC 131916

Qy    1177 attttcccctaccttgtaatccttgtgggattggaaaatagcttggtgatcacaaagagc 1236
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  131917 ATTTTCCCCTACCTTGTAATCCTTGTGGGATTGGAAAATAGCTTGGTGATCACAAAGAGC 131976

Qy    1237 gtagtctcaatggacgagacattcgacgtgaagatccgcgtggcgcaggctcttagcaag 1296
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  131977 GTAGTCTCAATGGACGAGACATTCGACGTGAAGATCCGCGTGGCGCAGGCTCTTAGCAAG 132036

Qy    1297 gagggttggcatatatccaagactcttttgacggagataacaattttgacaattggtctt 1356
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132037 GAGGGTTGGCATATATCCAAGACTCTTTTGACGGAGATAACAATTTTGACAATTGGTCTT 132096

Qy    1357 gctactttcgtgcccgtcatccaggagttttgtatctttgccatagtcggcttgctttcc 1416
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132097 GCTACTTTCGTGCCCGTCATCCAGGAGTTTTGTATCTTTGCCATAGTCGGCTTGCTTTCC 132156

Qy    1417 gattttatgctacagatgctgctcttctcaacaatactggccatgaacattaagcggacc 1476
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132157 GATTTTATGCTACAGATGCTGCTCTTCTCAACAATACTGGCCATGAACATTAAGCGGACC 132216

Qy    1477 gagtatacggcggaggccaagcaccttcctaagatgttgctgagctgcacccaaggggct 1536
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132217 GAGTATACGGCGGAGGCCAAGCACCTTCCTAAGATGTTGCTGAGCTGCACCCAAGGGGCT 132276

Qy    1537 ggtcgacaggatttccgatttttcggggccgccccagcactgccaccgtttgtccctggc 1596
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132277 GGTCGACAGGATTTCCGATTTTTCGGGGCCGCCCCAGCACTGCCACCGTTTGTCCCTGGC 132336

Qy    1597 acatttcagcgttctcagtcgcatccaaaactgtgttttgctgatcccgcatctgttagc 1656
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132337 ACATTTCAGCGTTCTCAGTCGCATCCAAAACTGTGTTTTGCTGATCCCGCATCTGTTAGC 132396

Qy    1657 gatcgtacaagcttggttaatggacactcgtcgccggagcaacgaatacccaaacgcata 1716
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132397 GATCGTACAAGCTTGGTTAATGGACACTCGTCGCCGGAGCAACGAATACCCAAACGCATA 132456

Qy    1717 aagattgtaaatttctgggcgcggactcgcttttttcagcgtgccttcatgatctggatg 1776
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132457 AAGATTGTAAATTTCTGGGCGCGGACTCGCTTTTTTCAGCGTGCCTTCATGATCTGGATG 132516

Qy    1777 attgtgtggatatgctctatagtttataattcgggatatctggagcagttgtttagcatg 1836
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132517 ATTGTGTGGATATGCTCTATAGTTTATAATTCGGGATATCTGGAGCAGTTGTTTAGCATG 132576

Qy    1837 cagagcaacggcacaatgacggcaacccttgaacttcaacggcgactacaggcgggtcgg 1896
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132577 CAGAGCAACGGCACAATGACGGCAACCCTTGAACTTCAACGGCGACTACAGGCGGGTCGG 132636

Qy    1897 ggagcagtcagcagttttttcgagggatggcaagcggacgggcagcgtgccacgagtgcg 1956
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132637 GGAGCAGTCAGCAGTTTTTTCGAGGGATGGCAAGCGGACGGGCAGCGTGCCACGAGTGCG 132696

Qy    1957 ccaagcggaagcggcttttctacgccaataaaagctcctctagcgatcgatataaacgaa 2016
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132697 CCAAGCGGAAGCGGCTTTTCTACGCCAATAAAAGCTCCTCTAGCGATCGATATAAACGAA 132756

Qy    2017 acggccgaggaaatgatgagacttcgatatcccagcttcgacctaaactatttccctttca 2076
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132757 ACGGCCGAGGAAATGATGAGACTTCGATATCCCAGCTTCGACCTAAACTATTTCCTTTCA 132816

Qy    2077 aacttccactggtccacgattatgaaacagtacaacatctcactaagtgggcactacgtt 2136
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132817 AACTTCCACTGGTCCACGATTATGAAACAGTACAACATCTCACTAAGTGGGCACTACGTT 132876

Qy    2137 accctgctaccgaccattcgccttagtcatgccatcgctccggagctagccactctgttg 2196
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132877 ACCCTGCTACCGACCATTCGCCTTAGTCATGCCATCGCTCCGGAGCTAGCCACTCTGTTG 132936

Qy    2197 cggaatccgcaggagcagctgcaacaaaatttttcaatggaaggccctagccgctgcactc 2256
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132937 CGGAATCCGCAGGAGCAGCTGCAACAAAATTTTCAATGGAAGGCCCTAGCCGCTGCACTC 132996

Qy    2257 gatccgctggactttaacgatgacgacgtgcgccgtgagtctccgatggtaatggcagag 2316
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132997 GATCCGCTGGACTTTAACGATGACGACGTGCGCCGTGAGTCTCCGATGGTAATGGCAGAG 133056

Qy    2317 gggttgcctctggttcccaagagccccatggaaatatttttcgccatcctcttgtgctgc 2376
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133057 GGGTTGCCTCTGGTTCCCAAGAGCCCCATGGAAATATTTTTCGCCATCCTCTTGTGCTGC 133116

Qy    2377 atcagcatcttcgtgctttgctacacgatggtggttttctaccgctgcatatgtaccagg 2436
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133117 ATCAGCATCTTCGTGCTTTGCTACACGATGGTGGTTTTCTACCGCTGCATATGTACCAGG 133176

Qy    2437 aactatgccgagtggcgctccagttggcacgaatctgaggcaccgtacaagcagactgag 2496
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133177 AACTATGCCGAGTGGCGCTCCAGTTGGCACGAATCTGAGGCACCGTACAAGCAGACTGAG 133236

Qy    2497 caaatcctggagggagttccaacgcaaatcgccggacacaaacatcgcattgaatgcctg 2556
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133237 CAAATCCTGGAGGGAGTTCCAACGCAAATCGCCGGACACAAACATCGCATTGAATGCCTG 133296

Qy    2557 gtgtctgacggcgcctacataatcagctgctgccttaaaggccaaatccgagtgtgggat 2616
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133297 GTGTCTGACGGCGCCTACATAATCAGCTGCTGCCTTAAAGGCCAAATCCGAGTGTGGGAT 133356

Qy    2617 gcacgcagtggcgagcagctaaccagcatctcccgatccgatattcagatctctcagcag 2676
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133357 GCACGCAGTGGCGAGCAGCTAACCAGCATCTCCCGATCCGATATTCAGATCTCTCAGCAG 133416

Qy    2677 cggacggatgggcagacgctggtacgaaagctggccgtgtcaccggtctggtgccttgac 2736
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133417 CGGACGGATGGGCAGACGCTGGTACGAAAGCTGGCCGTGTCACCGGTCTGGTGCCTTGAC 133476

Qy    2737 tacttcgataatctaatcgcagtaggctgcgccaacggccgcgtagaattgtgggaatcc 2796
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133477 TACTTCGATAATCTAATCGCAGTAGGCTGCGCCAACGGCCGCGTAGAATTGTGGGAATCC 133536

Qy    2797 cctgcgggattgcttaagtgtgcataccaggaagacgcgaagagaaaccagggtataacc 2856
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133537 CCTGCGGGATTGCTTAAGTGTGCATACCAGGAAGACGCGAAGAGAAACCAGGGTATAACC 133596

Qy    2857 cacatccacctgaacggcgatcgagtgattgtggcgcgtcttaatggccgactagatttt 2916
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133597 CACATCCACCTGAACGGCGATCGAGTGATTGTGGCGCGTCTTAATGGCCGACTAGATTTT 133656

Qy    2917 taccgcttagagacgtactacaaggggaagcaaatcgactggggttttacctcggcttac 2976
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db 133657 TACCGCTTAGAGACGTACTACAAGGGGAAGCAAATCGACTGGGGTTTTACCTCGGCTTAC 133716

Qy   2977 aggagaa---------------------------------------------------- 2983
          |||||||
Db 133717 AGGAGAAGTATGTCTTTCGCTTAAATACCCTTTTAAAAAACTCCTAAAAGAAAAGACTAT 133776

Qy   2984 ---------ctcatgttcgaactggatccactggaagcctgggattaatgttgcagcagc 3034
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 133777 TTCCTTCAGCTCATGTTCGAACTGGATCCACTGGAAGCCTGGGATTAATGTTGCAGCAGC 133836

Qy   3035 agcgctgtcagcaagaagcatcccagaagaccaccaaggaggaaatgaaaatcacattgg 3094
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 133837 AGCGCTGTCAGCAAGAAGCATCCCAGAAGACCACCAAGGAGGAAATGAAAATCACATTGG 133896

Qy   3095 agggtgtaagactagcccatcagcagccaatcacatgcatgcaggtcgttaacgacatgg 3154
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 133897 AGGGTGTAAGACTAGCCCATCAGCAGCCAATCACATGCATGCAGGTCGTTAACGACATGG 133956

Qy   3155 ttttcactggcagccaggatcacaccctca---------------------------- 3184
          ||||||||||||||||||||||||||||
Db 133957 TTTTCACTGGCAGCCAGGATCACACCCTCAAGGTAGGCATGAAACATTTTATTAAATCAA 134016

Qy   3185 -------------------------aggtgtattgcctcaataagtcggatgttgagta 3218
                                  ||||||||||||||||||||||||||||||||||
Db 134017 CTTTCCTTACTCTAACTGGTCTCCATAGGTGTATTGCCTCAATAAGTCGGATGTTGAGTA 134076

Qy   3219 tacgctccacggtcactgtgggcctgtaacctgtctctttgtggatcgctggcaacctgg 3278
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 134077 TACGCTCCACGGTCACTGTGGGCCTGTAACCTGTCTCTTTGTGGATCGCTGGCAACCTGG 134136

Qy   3279 cacaggggggtctgggtcccaggacggcctgctctgcgtatgggatctgttcacgggagc 3338
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 134137 CACAGGGGGGTCTGGGTCCCAGGACGGCCTGCTCTGCGTATGGGATCTGTTCACGGGAGC 134196

Qy   3339 ctgcatgtataatatacaagctcacgacggagccgtcagctgcctggcctgtgcgcccag 3398
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 134197 CTGCATGTATAATATACAAGCTCACGACGGAGCCGTCAGCTGCCTGGCCTGTGCGCCCAG 134256

Qy   3399 ttacgtaatctcgctaggcacggacgagaggatttgcgtatgggaacgatttcagggaaa 3458
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 134257 TTACGTAATCTCGCTAGGCACGGACGAGAGGATTTGCGTATGGGAACGATTTCAGGGAAA 134316

Qy   3459 cctgttgactaccatcaacatctcaaacgcatactcgagcctactgatgctaacaccgtc 3518
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 134317 CCTGTTGACTACCATCAACATCTCAAACGCATACTCGAGCCTACTGATGCTAACACCGTC 134376

Qy   3519 actattggttacgagcaaaatgggtaaggcctcattcttgattgccaatataagagggac 3578
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 134377 ACTATTGGTTACGAGCAAAATGGGTAAGGCCTCATTCTTGATTGCCAATATAAGAGGGAC 134436

Qy   3579 agtaaataataaatttaattccaacacaggatctcttattgtgtgggatgtgcgcactgg 3638
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 134437 AGTAAATAATAAATTTAATTCCAACACAGGATCTCTTATTGTGTGGGATGTGCGCACTGG 134496

Qy   3639 gcagccggctcgcgaggtcaaactggactttgcaaacctgcagctctgtcccaaaataat 3698
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 134497 GCAGCCGGCTCGCGAGGTCAAACTGGACTTTGCAAACCTGCAGCTCTGTCCCAAAATAAT 134556

Qy   3699 gatgcttgcctgcgattcggtagtttgcgactacggaaatgagatccgcgtcgtccgctt 3758
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 134557 GATGCTTGCCTGCGATTCGGTAGTTTGCGACTACGGAAATGAGATCCGCGTCGTCCGCTT 134616

Qy   3759 tcctatcgtggcagacaagtgccattaaagcgcaaaatttaatttagcgtggttcgcta 3818
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 134617 TCCTATCGTGGCAGACAAGTGCCATTAAAGCGCAAAATTTTAATTTAGCGTGGTTCGCTA 134676

Qy   3819 gcacctaggaataagttgacttaaggctttaaaacgcctggaagtcattgacgcattcac 3878
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 134677 GCACCTAGGAATAAGTTGACTTAAGGCTTTAAAACGCCTGGAAGTCATTGACGCATTCAC 134736

Qy   3879 tattttatataaatatatacactattagggtccgcagcaacttacggttttaacacaagc 3938
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 134737 TATTTTATATAAATATATACACTATTAGGGTCCGCAGCAACTTACGGTTTTAACACAAGC 134796

Qy   3939 tgtacgtatctcatctctagaattttgtgttagtttgtggacactaagtgtaacagctac 3998
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 134797 TGTACGTATCTCATCTCTAGAATTTTGTGTTAGTTTGTGGACACTAAGTGTAACAGCTAC 134856

Qy   3999 gctccggtaggttaaggaactaaactaaatgaatcagatatatacacatatattttcgcg 4058
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 134857 GCTCCGGTAGGTTAAGGAACTAAACTAAATGAATCAGATATATACACATATATTTTCGCG 134916

Qy   4059 taattatataaactacatagtgtcttaaagcgcctcagcctaatataaaatgactaaatg 4118
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 134917 TAATTATATAAACTACATAGTGTCTTAAAGCGCCTCAGCCTAATATAAAATGACTAAATG 134976

Qy   4119 ttaaaataaa 4128
          ||||||||||
Db 134977 TTAAAATAAA 134986


RESULT   3
AE003784
LOCUS       AE003784   313072 bp    DNA             INV       22-MAR-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386052 section 1
            of 5, complete sequence.
ACCESSION   AE003784 AE002769
VERSION     AE003784.1  GI:7302167
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 313072)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
```

```
REFERENCE   2  (bases 1 to 313072)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
FEATURES             Location/Qualifiers
     source          1. .313072
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
     mRNA            complement(join(1392. .1813,2404. .2648,2713. .3050,
                     3303. .4720,4806. .4989,5080. .5582,5806. .6319,6640. .7246,
                     13906. .14036,14558. .14629))
                     /gene="CG12110"
                     /product="CT40213"
                     /db_xref="FLYBASE:FBan0012110"
                     /db_xref="FLYBASE:FBgn0033075"
     mRNA            complement(join(1392. .1813,2404. .2648,2713. .3050,
                     3303. .4720,4806. .4989,5080. .5582,5806. .6319,6640. .7246,
                     13939. .14127,14558. .14634))
                     /gene="CG12110"
                     /product="CT40211"
                     /db_xref="FLYBASE:FBan0012110"
                     /db_xref="FLYBASE:FBgn0033075"
     gene            complement(<1392. .>14634)
                     /gene="CG12110"
                     /db_xref="FLYBASE:FBan0012110"
                     /db_xref="FLYBASE:FBgn0033075"
     CDS             complement(join(1712. .1813,2404. .2648,2713. .3050,
                     3303. .4720,4806. .4989,5080. .5582,5806. .6319,6640. .7139))
                     /gene="CG12110"
                     /note="CG12110 gene product [alt 1]"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0012110"
                     /db_xref="FLYBASE:FBgn0033075"
                     /protein_id="AAF57264.1"
                     /db_xref="GI:7302168"
                     /translation="MADAGTVGGALPPYALTNSDEEHGSGEEDASEENSNNEEGEGVF
                     RDCTDEAVVEHHNRCLPEFQFSLVDSEYDETLAFPDSVTILSNVGDKPVLVERKETDD
                     DEEEFDDEENNSVVLRHEIPFTSIYGPSVKFNSFQRKVFIPGREIHVRIVDTERSVTT
                     HLLNPNLYTIELTHGPFKWTIKRRYKHFNSLHQQLSFFRTSLNIPFPSRSHKEKRTTL
                     KATAREMADESTLKDLPSHTKVKQTSTPLRAEGRSSKIAGSNANNAMAMISPNHSSIL
                     AGLTPRRIQKKRKKKKKRKLPRFPNRPESLVTVENLSVRIKQLEDYLYNLLNISLYRS
                     HHETLNFVEVSNVSFVPGMGIKGKEGVILKRTGSTRPGQAGCNFFGCFQKNCCVRCNY
                     FCSDVVCGTWRNRWFFVKETCFGYIRPTDGSIRAVILFDQGFDVSTGIYQTGMRKGLQ
                     VLTNNRHIVLKCWTRRKCKEWMQYLKNTANSYARDFTLPNPHMSFAPMRANTHATCPE
                     IYMKRPALDGDYWRLDKILLRKAEQGVRVFVLLYKEVEMALGINSYYSKSTLAKHENI
                     KVMRHPDHARGGILLWAHHEKIVVIDQTYAFMGGIDLCYGRWDDHHHRLTDLGSISTS
                     SFSGSTRRTPSLYFTKDDTDSAFGSRKSSRNAHYDTSAKERPPSPPPDEPNTSIELKT
                     LKPGDRLLIPSTLVSSPGETPAESGIALEGMKLNTPEMERKNVLDRLKNNAMKGARMG
                     KDFMHRLTATETEEKSAEVYTIESEEATDHEVNLNMASGGQEVAITTSSTQILSEFCG
                     QAKYWFGKDYSNFILKDWMNLNSPFVDIIDRTTTPRMPWHDVGLCVVGTSARDVARHF
                     IQRWNAMKLEKLRDNTRFPYLMPKSYHQVRLNPNIQQNRQQRVTCQLLRSVSAWSCGF
                     IEADLVEQSIHDAYIQTITKAQHYVYIENQFFITMQLGMGVPGAYNNVRNQIGETLFK
                     RIVRAHKYETKILILILADLMFSSSRERKPFRVYVIMPLLPGFEGDVGGSTGIAVRAI
                     THWNYASISRGRTSILTRLQEAGIANPENYISFHSLRNHSFLNNTPITELIYVHSKLL
                     IADDRVVICGSANINDRSMIGKRDSEIAAILMDEEFEDGRMNGKKYPSGVFAGRLRKY
                     LFKEHLGLLESEGSSRSDLDINDPVCEKFWHGTWRRISMQNTEIYDEVFKCIPTDFVK
                     TFASLRKYQEEPPLAKTAPDLAANRANDIQGYLVDLPLEFLNKEVLTPPGTSKEGLIP
                     TSVWT"
     CDS             complement(join(1712. .1813,2404. .2648,2713. .3050,
                     3303. .4720,4806. .4989,5080. .5582,5806. .6319,6640. .7139))
                     /gene="CG12110"
                     /note="CG12110 gene product [alt 3]"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0012110"
                     /db_xref="FLYBASE:FBgn0033075"
                     /protein_id="AAF57265.1"
                     /db_xref="GI:7302169"
                     /translation="MADAGTVGGALPPYALTNSDEEHGSGEEDASEENSNNEEGEGVF
                     RDCTDEAVVEHHNRCLPEFQFSLVDSEYDETLAFPDSVTILSNVGDKPVLVERKETDD
                     DEEEFDDEENNSVVLRHEIPFTSIYGPSVKFNSFQRKVFIPGREIHVRIVDTERSVTT
                     HLLNPNLYTIELTHGPFKWTIKRRYKHFNSLHQQLSFFRTSLNIPFPSRSHKEKRTTL
                     KATAREMADESTLKDLPSHTKVKQTSTPLRAEGRSSKIAGSNANNAMAMISPNHSSIL
                     AGLTPRRIQKKRKKKKKRKLPRFPNRPESLVTVENLSVRIKQLEDYLYNLLNISLYRS
                     HHETLNFVEVSNVSFVPGMGIKGKEGVILKRTGSTRPGQAGCNFFGCFQKNCCVRCNY
                     FCSDVVCGTWRNRWFFVKETCFGYIRPTDGSIRAVILFDQGFDVSTGIYQTGMRKGLQ
                     VLTNNRHIVLKCWTRRKCKEWMQYLKNTANSYARDFTLPNPHMSFAPMRANTHATCPE
                     IYMKRPALDGDYWRLDKILLRKAEQGVRVFVLLYKEVEMALGINSYYSKSTLAKHENI
                     KVMRHPDHARGGILLWAHHEKIVVIDQTYAFMGGIDLCYGRWDDHHHRLTDLGSISTS
                     SFSGSTRRTPSLYFTKDDTDSAFGSRKSSRNAHYDTSAKERPPSPPPDEPNTSIELKT
                     LKPGDRLLIPSTLVSSPGETPAESGIALEGMKLNTPEMERKNVLDRLKNNAMKGARMG
                     KDFMHRLTATETEEKSAEVYTIESEEATDHEVNLNMASGGQEVAITTSSTQILSEFCG
                     QAKYWFGKDYSNFILKDWMNLNSPFVDIIDRTTTPRMPWHDVGLCVVGTSARDVARHF
                     IQRWNAMKLEKLRDNTRFPYLMPKSYHQVRLNPNIQQNRQQRVTCQLLRSVSAWSCGF
                     IEADLVEQSIHDAYIQTITKAQHYVYIENQFFITMQLGMGVPGAYNNVRNQIGETLFK
                     RIVRAHKYETKILILILADLMFSSSRERKPFRVYVIMPLLPGFEGDVGGSTGIAVRAI
                     THWNYASISRGRTSILTRLQEAGIANPENYISFHSLRNHSFLNNTPITELIYVHSKLL
                     IADDRVVICGSANINDRSMIGKRDSEIAAILMDEEFEDGRMNGKKYPSGVFAGRLRKY
                     LFKEHLGLLESEGSSRSDLDINDPVCEKFWHGTWRRISMQNTEIYDEVFKCIPTDFVK
                     TFASLRKYQEEPPLAKTAPDLAANRANDIQGYLVDLPLEFLNKEVLTPPGTSKEGLIP
                     TSVWT"
     mRNA            complement(join(2380. .2540,6640. .7246,13939. .14624))
                     /gene="CG12110"
                     /product="CT6556"
                     /db_xref="FLYBASE:FBan0012110"
                     /db_xref="FLYBASE:FBgn0033075"
     CDS             complement(join(2519. .2540,6640. .7139))
                     /gene="CG12110"
                     /note="CG12110 gene product [alt 2]"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0012110"
                     /db_xref="FLYBASE:FBgn0033075"
                     /protein_id="AAF57266.1"
                     /db_xref="GI:7302170"
                     /translation="MADAGTVGGALPPYALTNSDEEHGSGEEDASEENSNNEEGEGVF
                     RDCTDEAVVEHHNRCLPEFQFSLVDSEYDETLAFPDSVTILSNVGDKPVLVERKETDD
                     DEEEFDDEENNSVVLRHEIPFTSIYGPSVKFNSFQRKVFIPGREIHVRIVDTERSVTT
                     HLLNPNLDLRRGV"
     mRNA            complement(join(16896. .17675,17748. .17894,17986. .19055,
                     19117. .19315,20599. .21889,22351. .22643,22718. .23392,
                     35382. .35595))
                     /gene="CG8276"
                     /note="Transcript structure is suspect."
                     /product="CT24495"
                     /db_xref="FLYBASE:FBan0008276"
                     /db_xref="FLYBASE:FBgn0033073"
     gene            complement(<16896. .>35595)
                     /gene="CG8276"
                     /db_xref="FLYBASE:FBan0008276"
                     /db_xref="FLYBASE:FBgn0033073"
     CDS             complement(join(22552. .22643,22718. .22967))
                     /gene="CG8276"
                     /note="CG8276 gene product; Transcript structure is
                     suspect"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0008276"
                     /db_xref="FLYBASE:FBgn0033073"
                     /protein_id="AAF57267.1"
                     /db_xref="GI:7302171"
                     /translation="MEKRLSDSPGDCRVTRSTMTPTLRLDQTSRQEPLPQQPDNGPAA
                     APGKSKSPTPLPGKSQAAQHHQFRAPQQQQGPKNRNKAWIYGLLGHSRDSPSHCGVGS
                     QGGSREHPEYP"
     mRNA            complement(<29911. .>30933)
                     /gene="CG8330"
                     /product="CT4956"
                     /db_xref="FLYBASE:FBan0008330"
                     /db_xref="FLYBASE:FBgn0033074"
     gene            complement(<29911. .>30933)
                     /gene="CG8330"
                     /db_xref="FLYBASE:FBan0008330"
                     /db_xref="FLYBASE:FBgn0033074"
     CDS             complement(29911. .30933)
                     /gene="CG8330"
                     /note="CG8330 gene product"
                     /codon_start=1
```

/db_xref="FLYBASE:FBan0008330"

Query Match 79.4%; Score 3279.6; DB 58; Length 313072;
Best Local Similarity 96.4%; Pred. No. 0;
Matches 3423; Conservative 0; Mismatches 9; Indels 118; Gaps 2;

```
Qy    697 acgatatttcagtaccataacctacagaaatcaaaagtttccgcggcggaaatgctgttt 756
          | | |  | || |  || ||||||||||||||||||||||||||||||||||||||||||||
Db 226681 ATGTTTGTCCATTTTCAGAACCTACAGAAATCAAAAGTTTCCGCGGCGGAAATGCTGTTT 226740

Qy    757 ggattacccatgcaggacactggattcaagcgctatccattgcgcgctcggtcgcgtatt 816
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 226741 GGATTACCCATGCAGGACACTGGATTCAAGCGCTATCCATTGCGCGCTCGGTCGCGTATT 226800

Qy    817 atacagtatgccttgacgttattcctcaagcacaacgatatggagtatctggacactcta 876
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 226801 ATACAGTATGCCTTGACGTTATTCCTCAAGCACAACGATATGGAGTATCTGGACACTCTA 226860

Qy    877 aaggaaaaagctgctgcgacactatcccccactcccgttggctagtgcgtcggctgaagag 936
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 226861 AAGGAAAAGCTGCTGCGACACTATCCCCCACTCCCGTTGGCTAGTGCGTCGGCTGAAGAG 226920

Qy    937 ccgacgaccataacttacatcttttatccaggagagtacaggatgtgggagctggtgcct 996
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 226921 CCGACGACCATAACTTACATCTTTTATCCAGGAGAGTACAGGATGTGGGAGCTGGTGCCT 226980

Qy    997 tacacagtggcctttatgttggtgtttgcttatgtgtacttctctgttcgaaaaatcgat 1056
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 226981 TACACAGTGGCCTTTATGTTGGTGTTTGCTTATGTGTACTTCTCTGTTCGAAAAATCGAT 227040

Qy   1057 gtatttcgttcccgcttttgctggccttatgtagcgtaatcaccacagccgggagcttg 1116
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 227041 GTATTTCGTTCCCGCTTTTGCTGGCCTTATGTAGCGTAATCACCACAGCCGGGAGCTTG 227100

Qy   1117 gccatgtcccttggcttgtgtttcttctttggcctgacaatttcgctgcagtcaaaggac 1176
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 227101 GCCATGTCCCTTGGCTTGTGTTTCTTCTTTGGCCTGACAATTTCGCTGCAGTCAAAGGAC 227160

Qy   1177 attttcccctaccttgtaatccttgtgggattggaaaatagcttggtgatcacaaagagc 1236
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 227161 ATTTTCCCCTACCTTGTAATCCTTGTGGGATTGGAAAATAGCTTGGTGATCACAAAGAGC 227220

Qy   1237 gtagtctcaatggacgagacattcgacgtgaagatccgcgtggcgcaggctcttagcaag 1296
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 227221 GTAGTCTCAATGGACGAGACATTCGACGTGAAGATCCGCGTGGCGCAGGCTCTTAGCAAG 227280

Qy   1297 gagggttggcatatatccaagactcttttgacggagataacaattttgacaattggtctt 1356
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 227281 GAGGGTTGGCATATATCCAAGACTCTTTTGACGGAGATAACAATTTTGACAATTGGTCTT 227340

Qy   1357 gctactttcgtgcccgtcatccaggagttttgtatctttgccatagtcggcttgctttcc 1416
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 227341 GCTACTTTCGTGCCCGTCATCCAGGAGTTTTGTATCTTTGCCATAGTCGGCTTGCTTTCC 227400

Qy   1417 gattttatgctacagatgctgctcttctcaacaatactggccatgaacattaagcggacc 1476
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 227401 GATTTTATGCTACAGATGCTGCTCTTCTCAACAATACTGGCCATGAACATTAAGCGGACC 227460

Qy   1477 gagtatacggcggaggccaagcaccttcctaagatgttgctgagctgcacccaaggggct 1536
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 227461 GAGTATACGGCGGAGGCCAAGCACCTTCCTAAGATGTTGCTGAGCTGCACCCAAGGGGCT 227520

Qy   1537 ggtcgacaggatttccgatttttcggggccgccccagcactgccaccgtttgtccctggc 1596
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 227521 GGTCGACAGGATTTCCGATTTTTCGGGGCCGCCCCAGCACTGCCACCGTTTGTCCCTGGC 227580

Qy   1597 acatttcagcgttctcagtcgcatccaaaactgtgttttgctgatcccgcatctgttagc 1656
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 227581 ACATTTCAGCGTTCTCAGTCGCATCCAAAACTGTGTTTTGCTGATCCCGCATCTGTTAGC 227640

Qy   1657 gatcgtacaagcttggttaatggacactcgtcgccggagcaacgaatacccaaacgcata 1716
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 227641 GATCGTACAAGCTTGGTTAATGGACACTCGTCGCCGGAGCAACGAATACCCAAACGCATA 227700

Qy   1717 aagattgtaaatttctgggcgcggactcgctttttcagcgtgccttcatgatctggatg 1776
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 227701 AAGATTGTAAATTTCTGGGCGCGGACTCGCTTTTTCAGCGTGCCTTCATGATCTGGATG 227760

Qy   1777 attgtgtggatatgctctatagtttataattcgggatatctggagcagttgtttagcatg 1836
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 227761 ATTGTGTGGATATGCTCTATAGTTTATAATTCGGGATATCTGGAGCAGTTGTTTAGCATG 227820

Qy   1837 cagagcaacggcacaatgacggcaacccttgaacttcaacggcgactacaggcgggtcgg 1896
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 227821 CAGAGCAACGGCACAATGACGGCAACCCTTGAACTTCAACGGCGACTACAGGCGGGTCGG 227880

Qy   1897 ggagcagtcagcagtttttcgagggatggcaagcggacgggcagcgtgccacgagtgcg 1956
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 227881 GGAGCAGTCAGCAGTTTTTCGAGGGATGGCAAGCGGACGGGCAGCGTGCCACGAGTGCG 227940

Qy   1957 ccaagcggaagcggctttttctacgccaataaaagctcctctagcgatcgatataaacgaa 2016
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 227941 CCAAGCGGAAGCGGCTTTTCTACGCCAATAAAAGCTCCTCTAGCGATCGATATAAACGAA 228000

Qy   2017 acggccgaggaaatgatgagacttcgatatcccagcttcgacctaaactatttccttttca 2076
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228001 ACGGCCGAGGAAATGATGAGACTTCGATATCCCAGCTTCGACCTAAACTATTTCCTTTCA 228060

Qy   2077 aacttccactggtccacgattatgaaacagtacaacatctcactaagtgggcactacgtt 2136
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228061 AACTTCCACTGGTCCACGATTATGAAACAGTACAACATCTCACTAAGTGGGCACTACGTT 228120

Qy   2137 accctgctaccgaccattcgccttagtcatgccatcgctccggagctagccactctgttg 2196
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228121 ACCCTGCTACCGACCATTCGCCTTAGTCATGCCATCGCTCCGGAGCTAGCCACTCTGTTG 228180

Qy   2197 cggaatccgcaggagcagctgcaacaaaattttcaatggaaggccctagccgctgcactc 2256
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228181 CGGAATCCGCAGGAGCAGCTGCAACAAAATTTTCAATGGAAGGCCCTAGCCGCTGCACTC 228240

Qy   2257 gatccgctggactttaacgatgacgacgtgcgccgtgagtctccgatggtaatggcagag 2316
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228241 GATCCGCTGGACTTTAACGATGACGACGTGCGCCGTGAGTCTCCGATGGTAATGGCAGAG 228300

Qy   2317 gggttgcctctggttcccaagagccccatggaaatatttttcgccatcctcttgtgctgc 2376
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228301 GGGTTGCCTCTGGTTCCCAAGAGCCCCATGGAAATATTTTTCGCCATCCTCTTGTGCTGC 228360

Qy   2377 atcagcatcttcgtgctttgctacacgatggtggttttctaccgctgcatatgtaccagg 2436
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228361 ATCAGCATCTTCGTGCTTTGCTACACGATGGTGGTTTTCTACCGCTGCATATGTACCAGG 228420

Qy   2437 aactatgccgagtggcgctccagttggcacgaatctgaggcaccgtacaagcagactgag 2496
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228421 AACTATGCCGAGTGGCGCTCCAGTTGGCACGAATCTGAGGCACCGTACAAGCAGACTGAG 228480

Qy   2497 caaatcctggagggagttccaacgcaaatcgccggacacaaacatcgcattgaatgcctg 2556
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228481 CAAATCCTGGAGGGAGTTCCAACGCAAATCGCCGGACACAAACATCGCATTGAATGCCTG 228540

Qy   2557 gtgtctgacggcgcctacataatcagctgctgccttaaaggccaaatccgagtgtgggat 2616
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228541 GTGTCTGACGGCGCCTACATAATCAGCTGCTGCCTTAAAGGCCAAATCCGAGTGTGGGAT 228600

Qy   2617 gcacgcagtggcgagcagctaaccagcatctcccgatccgatattcagatctctcagcag 2676
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228601 GCACGCAGTGGCGAGCAGCTAACCAGCATCTCCCGATCCGATATTCAGATCTCTCAGCAG 228660

Qy   2677 cggacggatgggcagacgctggtacgaaagctggccgtgtcaccggtctggtgccttgac 2736
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228661 CGGACGGATGGGCAGACGCTGGTACGAAAGCTGGCCGTGTCACCGGTCTGGTGCCTTGAC 228720

Qy   2737 tacttcgataatctaatcgcagtaggctgcgccaacggccgcgtagaattgtgggaatcc 2796
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228721 TACTTCGATAATCTAATCGCAGTAGGCTGCGCCAACGGCCGCGTAGAATTGTGGGAATCC 228780
```

```
Qy   2797 cctgcgggattgcttaagtgtgcataccaggaagacgcgaagagaaaccagggtataacc 2856
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228781 CCTGCGGGATTGCTTAAGTGTGCATACCAGGAAGACGCGAAGAGAAACCAGGGTATAACC 228840

Qy   2857 cacatccacctgaacggcgatcgagtgattgtggcgcgtcttaatggccgactagatttt 2916
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228841 CACATCCACCTGAACGGCGATCGAGTGATTGTGGCGCGTCTTAATGGCCGACTAGATTTT 228900

Qy   2917 taccgcttagagacgtactacaaggggaagcaaatcgactggggttttacctcggcttac 2976
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228901 TACCGCTTAGAGACGTACTACAAGGGGAAGCAAATCGACTGGGGTTTTACCTCGGCTTAC 228960

Qy   2977 aggagaa------------------------------------------------------ 2983
          |||||||
Db 228961 AGGAGAAGTATGTCTTTCGCTTAAATACCCTTTTAAAAAACTCCTAAAAGAAAAGACTAT 229020

Qy   2984 ---------ctcatgttcgaactggatccactggaagcctgggattaatgttgcagcagc 3034
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229021 TTCCTTCAGCTCATGTTCGAACTGGATCCACTGGAAGCCTGGGATTAATGTTGCAGCAGC 229080

Qy   3035 agcgctgtcagcaagaagcatcccagaagaccaccaaggaggaaatgaaaatcacattgg 3094
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229081 AGCGCTGTCAGCAAGAAGCATCCCAGAAGACCACCAAGGAGGAAATGAAAATCACATTGG 229140

Qy   3095 agggtgtaagactagcccatcagcagccaatcacatgcatgcaggtcgttaacgacatgg 3154
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229141 AGGGTGTAAGACTAGCCCATCAGCAGCCAATCACATGCATGCAGGTCGTTAACGACATGG 229200

Qy   3155 ttttcactggcagccaggatcacaccctca------------------------------ 3184
          ||||||||||||||||||||||||||||||
Db 229201 TTTTCACTGGCAGCCAGGATCACACCCTCAAGGTAGGCATGAAACATTTTATTAAATCAA 229260

Qy   3185 --------------------------aggtgtattgcctcaataagtcggatgttgagta 3218
                                   ||||||||||||||||||||||||||||||||||
Db 229261 CTTTCCTTACTCTAACTGGTCTCCATAGGTGTATTGCCTCAATAAGTCGGATGTTGAGTA 229320

Qy   3219 tacgctccacggtcactgtgggcctgtaacctgtctctttgtggatcgctggcaacctgg 3278
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229321 TACGCTCCACGGTCACTGTGGGCCTGTAACCTGTCTCTTTGTGGATCGCTGGCAACCTGG 229380

Qy   3279 cacaggggggtctgggtcccaggacggcctgctctgcgtatgggatctgttcacgggagc 3338
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229381 CACAGGGGGGTCTGGGTCCCAGGACGGCCTGCTCTGCGTATGGGATCTGTTCACGGGAGC 229440

Qy   3339 ctgcatgtataatatacaagctcacgacggagccgtcagctgcctggcctgtgcgcccag 3398
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229441 CTGCATGTATAATATACAAGCTCACGACGGAGCCGTCAGCTGCCTGGCCTGTGCGCCCAG 229500

Qy   3399 ttacgtaatctcgctaggcacggacgagaggatttgcgtatgggaacgatttcagggaaa 3458
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229501 TTACGTAATCTCGCTAGGCACGGACGAGAGGATTTGCGTATGGGAACGATTTCAGGGAAA 229560

Qy   3459 cctgttgactaccatcaacatctcaaacgcatactcgagcctactgatgctaacaccgtc 3518
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229561 CCTGTTGACTACCATCAACATCTCAAACGCATACTCGAGCCTACTGATGCTAACACCGTC 229620

Qy   3519 actattggttacgagcaaaatgggtaaggcctcattcttgattgccaatataagagggac 3578
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229621 ACTATTGGTTACGAGCAAAATGGGTAAGGCCTCATTCTTGATTGCCAATATAAGAGGGAC 229680

Qy   3579 agtaaataataaatttaattccaacacaggatctcttattgtgtgggatgtgcgcactgg 3638
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229681 AGTAAATAATAAATTTAATTCCAACACAGGATCTCTTATTGTGTGGGATGTGCGCACTGG 229740

Qy   3639 gcagccggctcgcgaggtcaaactggactttgcaaacctgcagctctgtcccaaaataat 3698
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229741 GCAGCCGGCTCGCGAGGTCAAACTGGACTTTGCAAACCTGCAGCTCTGTCCCAAAATAAT 229800

Qy   3699 gatgcttgcctgcgattcggtagtttgcgactacggaaatgagatccgcgtcgtccgctt 3758
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229801 GATGCTTGCCTGCGATTCGGTAGTTTGCGACTACGGAAATGAGATCCGCGTCGTCCGCTT 229860

Qy   3759 tcctatcgtggcagacaagtgccattaaagcgcaaaattttaatttagcgtggttcgcta 3818
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229861 TCCTATCGTGGCAGACAAGTGCCATTAAAGCGCAAAATTTTAATTTAGCGTGGTTCGCTA 229920

Qy   3819 gcacctaggaataagttgacttaaggctttaaaacgcctggaagtcattgacgcattcac 3878
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229921 GCACCTAGGAATAAGTTGACTTAAGGCTTTAAAACGCCTGGAAGTCATTGACGCATTCAC 229980

Qy   3879 tattttatataaatatatacactattagggtccgcagcaacttacggttttaacacaagc 3938
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229981 TATTTTATATAAATATATACACTATTAGGGTCCGCAGCAACTTACGGTTTTAACACAAGC 230040

Qy   3939 tgtacgtatctcatctctagaattttgtgttagtttgtggacactaagtgtaacagctac 3998
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 230041 TGTACGTATCTCATCTCTAGAATTTTGTGTTAGTTTGTGGACACTAAGTGTAACAGCTAC 230100

Qy   3999 gctccggtaggttaaggaactaaactaaatgaatcagatatatacacatatattttcgcg 4058
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 230101 GCTCCGGTAGGTTAAGGAACTAAACTAAATGAATCAGATATATACACATATATTTTCGCG 230160

Qy   4059 taattatataaactacatagtgtcttaaagcgcctcagcctaatataaaatgactaaatg 4118
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 230161 TAATTATATAAACTACATAGTGTCTTAAAGCGCCTCAGCCTAATATAAAATGACTAAATG 230220

Qy   4119 ttaaaataaa 4128
          ||||||||||
Db 230221 TTAAAATAAA 230230


RESULT   4
AC020090
LOCUS       AC020090    60625 bp    DNA             HTG       03-JAN-2000
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC020090
VERSION     AC020090.1  GI:6664807
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 60625)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212008 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .60625
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    17399 a  13001 c  12988 g  17237 t
ORIGIN


  Query Match             57.9%;  Score 2390.4;  DB 43;  Length 60625;
  Best Local Similarity   97.1%;  Pred. No. 0;
  Matches 2480;  Conservative    0;  Mismatches   11;  Indels   63;  Gaps    2;

Qy    697 acgatatttcagtaccataacctacagaaatcaaaagtttccgcggcggaaatgctgttt 756
          | | |  | || |  || ||||||||||||||||||||||||||||||||||||||||||||
Db  57951 ATGTTTGTCCATTTTCAGAACCTACAGAAATCAAAAGTTTCCGCGGCGGAAATGCTGTTT 58010

Qy    757 ggattacccatgcaggacactggattcaagcgctatccattgcgcgctcggtcgcgtatt 816
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  58011 GGATTACCCATGCAGGACACTGGATTCAAGCGCTATCCATTGCGCGCTCGGTCGCGTATT 58070

Qy    817 atacagtatgccttgacgttattcctcaagcacaacgatatggagtatctggacactcta 876
```

```
Db  58071 ATACAGTATG-CTTGACGTTATTCCTCAAGCACAACGATATGGAGTATCTGGACACTCTA 58129

Qy    877 aaggaaaagctgctgcgacactatcccccactcccgttggctagtgcgtcggctgaagag 936

Db  58130 AAGGAAAAGCTGCTGCGACACTATCCCCCACTCCCGTTGGCTAGTGCGTCGGCTGAAGAG 58189

Qy    937 ccgacgaccataacttacatcttttatccaggagagtacaggatgtgggagctggtgcct 996

Db  58190 CCGACGACCATAACTTACATCTTTTATCCAGGAGAGTACAGGATGTGGGAGCTGGTGCCT 58249

Qy    997 tacacagtggcctttatgttggtgtttgcttatgtgtacttctctgttcgaaaaatcgat 1056

Db  58250 TACACAGTGGCCTTTATGTTGGTGTTTGCTTATGTGTACTTCTCTGTTCGAAAAATCGAT 58309

Qy   1057 gtatttcgttcccgcttttgctggccttatgtagcgtaatcaccacagccgggagcttg 1116

Db  58310 GTATTTCGTTCCCGCTTTTTGCTGGCCTTATGTAGCGTAATCACCACAGCCGGGAGCTTG 58369

Qy   1117 gccatgtcccttggcttgtgtttcttctttggcctgacaatttcgctgcagtcaaaggac 1176

Db  58370 GCCATGTCCCTTGGCTTGTGTTTCTTCTTTGGCCTGACAATTTCGCTGCAGTCAAAGGAC 58429

Qy   1177 attttcccctaccttgtaatccttgtgggattggaaaatagcttggtgatcacaaagagc 1236

Db  58430 ATTTTCCCCTACCTTGTAATCCTTGTGGGATTGGAAAATAGCTTGGTGATCACAAAGAGC 58489

Qy   1237 gtagtctcaatggacgagacattcgacgtgaagatccgcgtggcgcaggctcttagcaag 1296

Db  58490 GTAGTCTCAATGGACGAGACATTCGACGTGAAGATCCGCGTGGCGCAGGCTCTTAGCAAG 58549

Qy   1297 gagggttggcatatatccaagactcttttgacggagataacaattttgacaattggtctt 1356

Db  58550 GAGGGTTGGCATATATCCAAGACTCTTTTGACGGAGATAACAATTTTGACAATTGGTCTT 58609

Qy   1357 gctactttcgtgcccgtcatccaggagttttgtatctttgccatagtcggcttgctttcc 1416

Db  58610 GCTACTTTCGTGCCCGTCATCCAGGAGTTTTGTATCTTTGCCATAGTCGGCTTGCTTTCC 58669

Qy   1417 gattttatgctacagatgctgctcttctcaacaatactggccatgaacattaagcggacc 1476

Db  58670 GATTTTATGCTACAGATGCTGCTCTACTCAACAATACTGGCCATGAACATTAAGCGGACC 58729

Qy   1477 gagtatacggcggaggccaagcaccttcctaagatgttgctgagctgcacccaaggggct 1536

Db  58730 GAGTATACGGCGGAGGCCAAGCACCTTCCTAAGATGTTGCTGAGCTGCACCCAAGGGGCT 58789

Qy   1537 ggtcgacaggatttccgatttttcggggccgccccagcactgccaccgtttgtccctggc 1596

Db  58790 GGTCGACAGGATTTCCGATTTTTCGGGGCCGCCCCAGCACTGCCACCGTTTGTCCCTGGC 58849

Qy   1597 acatttcagcgttctcagtcgcatccaaaactgtgttttgctgatcccgcatctgttagc 1656

Db  58850 ACATTTCAGCGTTCTCAGTCGCATCCAAAACTGTGTTTTGCTGATCCCGCATCTGTTAGC 58909

Qy   1657 gatcgtacaagcttggttaatggacactcgtcgccggagcaacgaatacccaaacgcata 1716

Db  58910 GATCGTACAAGCTTGGTTAATGGACACTCGTCGCCGGAGCAACGAATACCCAAACGCATA 58969

Qy   1717 aagattgtaaatttctgggcgcggactcgctttttttcagcgtgccttcatgatctggatg 1776

Db  58970 AAGATTGTAAATTTCTGGGCGCGGACTCGCTTTTTTCAGCGTGCCTTCATGATCTGGATG 59029

Qy   1777 attgtgtggatatgctctatagtttataattcgggatatctggagcagttgtttagcatg 1836

Db  59030 ATTGTGTGGATATGCTCTATAGTTTATAATTCGGGATATCTGGAGCAGTTGTTTAGCATG 59089

Qy   1837 cagagcaacggcacaatgacggcaacccttgaacttcaacggcgactacaggcgggtcgg 1896

Db  59090 CAGAGCAACGGCACAATGACGGCAACCCTTGAACTTCAACGGCGACTACAGGCGGGTCGG 59149

Qy   1897 ggagcagtcagcagttttttcgagggatggcaagcggacgggcagcgtgccacgagtgcg 1956

Db  59150 GGAGCAGTCAGCAGTTTTTTCGAGGGATGGCAAGCGGACGGGCAGCGTGCCACGAGTGCG 59209

Qy   1957 ccaagcggaagcggcttttctacgccaataaaagctcctctagcgatcgatataaacgaa 2016

Db  59210 CCAAGCGGAAGCGGCTTTTCTACGCCAATAAAAGCTCCTCTAGCGATCGATATAAACGAA 59269

Qy   2017 acggccgaggaaatgatgagacttcgatatcccagcttcgacctaaactatttccttttca 2076

Db  59270 ACGGCCGAGGAAATGATGAGACTTCGATATCCCAGCTTCGACCTAAACTATTTCCTTTCA 59329

Qy   2077 aacttccactggtccacgattatgaaacagtacaacatctcactaagtgggcactacgtt 2136

Db  59330 AACTTCCACTGGTCCACGATTATGAAACAGTACAACATCTCACTAAGTGGGCACTACGTT 59389

Qy   2137 accctgctaccgaccattcgccttagtcatgccatcgctccggagctagccactctgttg 2196

Db  59390 ACCCTGCTACCGACCATTCGCCTTAGTCATGCCATCGCTCCGGAGCTAGCCACTCTGTTG 59449

Qy   2197 cggaatccgcaggagcagctgcaacaaaatttttcaatggaaggccctagccgctgcactc 2256

Db  59450 CGGAATCCGCAGGAGCAGCTGCAACAAAATTTTCAATGGAAGGCCCTAGCCGCTGCACTC 59509

Qy   2257 gatccgctggactttaacgatgacgacgtgcgccgtgagtctccgatggtaatggcagag 2316

Db  59510 GATCCGCTGGACTTTAACGATGACGACGTGCGCCGTGAGTCTCCGATGGTAATGGCAGAG 59569

Qy   2317 gggttgcctctggttcccaagagccccatggaaatatttttcgccatcctcttgtgctgc 2376

Db  59570 GGGTTGCCTCTGGTTCCCAAGAGCCCCATGGAAATATTTTTCGCCATCCTCTTGTGCTGC 59629

Qy   2377 atcagcatcttcgtgctttgctacacgatggtggttttctaccgctgcatatgtaccagg 2436

Db  59630 ATCAGCATCTTCGTGCTTTGCTACACGATGGTGGTTTTCTACCGCTGCATATGTACCAGG 59689

Qy   2437 aactatgccgagtggcgctccagttggcacgaatctgaggcaccgtacaagcagactgag 2496

Db  59690 AACTATGCCGAGTGGCGCTCCAGTTGGCACGAATCTGAGGCACCGTACAAGCAGACTGAG 59749

Qy   2497 caaatcctggagggagttccaacgcaaatcgccggacacaaacatcgcattgaatgcctg 2556

Db  59750 CAAATCCTGGAGGGAGTTCCAACGCAAATCGCCGGACACAAACATCGCATTGAATGCCTG 59809

Qy   2557 gtgtctgacggcgcctacataatcagctgctgccttaaaggccaaatccgagtgtgggat 2616

Db  59810 GTGTCTGACGGCGCCTACATAATCAGCTGCTGCCTTAAAGGCCAAATCCGAGTGTGGGAT 59869

Qy   2617 gcacgcagtggcgagcagctaaccagcatctcccgatccgatattcagatctctcagcag 2676

Db  59870 GCACGCAGTGGCGAGCAGCTAACCAGCATCTCCCGATCCGATATTCAGATCTCTCAGCAG 59929

Qy   2677 cggacggatgggcagacgctggtacgaaagctggccgtgtcaccggtctggtgccttgac 2736

Db  59930 CGGACGGATGGGCAGACGCTGGTACGAAAGCTGGCCGTGTCACCGGTCTGGTGCCTTGAC 59989

Qy   2737 tacttcgataatctaatcgcagtaggctgcgccaacggccgcgtagaattgtgggaatcc 2796

Db  59990 TACTTCGATAATCTAATCGCAGTAGGCTGCGCCAACGGCCGCGTAGAATTGTGGGAATCC 60049

Qy   2797 cctgcgggattgcttaagtgtgcataccaggaagacgcgaagagaaaccagggtataacc 2856

Db  60050 CCTGCGGGATTGCTTAAGTGTGCATACCAGGAAGACGCGAAGAGAAACCAGGGTATAACC 60109

Qy   2857 cacatccacctgaacggcgatcgagtgattgtggcgcgtcttaatggccgactagatttt 2916

Db  60110 CACATCCACCTGAACGGCGATCGAGTGATTGTGGCGCGTCTTAATGGCCGACTAGATTTT 60169

Qy   2917 taccgcttagagacgtactacaaggggaagcaaatcgactggggttttacctcggcttac 2976

Db  60170 TACCGCTTAGAGACGTACTACAAGGGGAAGCAAATCGACTGGGGTTTTACCTCGGCTTAC 60229

Qy   2977 aggagaa------------------------------------------------------ 2983

Db  60230 AGGAGAAGTATGTCTTTCGCTTAAATACCCTTTTAAAAAACTCCTAAAAGAAAAGACTAT 60289
```

```
Qy   2984 ---------ctcatgttcgaactggatccactggaagcctgggattaatgttgcagcagc 3034
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  60290 TTCCTTCAGCTCATGTTCGAACTGGATCCACTGGAAGCCTGGGATTAATGTTGCAGCAGC 60349

Qy   3035 agcgctgtcagcaagaagcatcccagaagaccaccaaggaggaaatgaaaatcacattgg 3094
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db  60350 AGCGCTGTCAGCAAGAAGCATCCCAGTAGACCACCAAGGAGGAAATGAAAATCACATTGG 60409

Qy   3095 agggtgtaagactagcccatcagcagccaatcacatgcatgcaggtcgttaacgacatgg 3154
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  60410 AGGGTGTAAGACTAGCCCATCAGCAGCCAATCACATGCATGCAGGTCGTTAACGACATGG 60469

Qy   3155 ttttcactggcagccaggatcacaccctcaaggt 3188
          |||||||||||||||||||||||||||||||||
Db  60470 TTTTCACTGGCAGCCAGGATCACACCCTCAAGGT 60503


RESULT   5
AC014229
LOCUS       AC014229    12978 bp    DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC014229
VERSION     AC014229.1  GI:6437106
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 12978)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212393 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .12978
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     3902 a   2637 c   2629 g   3810 t
ORIGIN


  Query Match             27.3%;  Score 1126.4;  DB 42;  Length 12978;
  Best Local Similarity   91.4%;  Pred. No. 0;
  Matches 1265;  Conservative    0;  Mismatches    1;  Indels  118;  Gaps    2;

Qy   2863 cacctgaacggcgatcgagtgattgtggcgcgtcttaatggccgactagatttttaccgc 2922
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11 CCCCTGAACGGCGATCGAGTGATTGTGGCGCGTCTTAATGGCCGACTAGATTTTTACCGC 70

Qy   2923 ttagagacgtactacaaggggaagcaaatcgactggggttttacctcggcttacaggaga 2982
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     71 TTAGAGACGTACTACAAGGGGAAGCAAATCGACTGGGGTTTTACCTCGGCTTACAGGAGA 130

Qy   2983 a---------------------------------------------------------- 2983
           |
Db    131 AGTATGTCTTTCGCTTAAATACCCTTTTAAAAAACTCCTAAAAGAAAAGACTATTTCCTT 190

Qy   2984 ---ctcatgttcgaactggatccactggaagcctgggattaatgttgcagcagcagcgct 3040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    191 CAGCTCATGTTCGAACTGGATCCACTGGAAGCCTGGGATTAATGTTGCAGCAGCAGCGCT 250

Qy   3041 gtcagcaagaagcatcccagaagaccaccaaggaggaaatgaaaatcacattggagggtg 3100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    251 GTCAGCAAGAAGCATCCCAGAAGACCACCAAGGAGGAAATGAAAATCACATTGGAGGGTG 310

Qy   3101 taagactagcccatcagcagccaatcacatgcatgcaggtcgttaacgacatggttttca 3160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    311 TAAGACTAGCCCATCAGCAGCCAATCACATGCATGCAGGTCGTTAACGACATGGTTTTCA 370

Qy   3161 ctggcagccaggatcacaccctca---------------------------------- 3184
          |||||||||||||||||||||||
Db    371 CTGGCAGCCAGGATCACACCCTCAAGGTAGGCATGAAACATTTTATTAAATCAACTTTCC 430

Qy   3185 --------------------aggtgtattgcctcaataagtcggatgttgagtatacgct 3224
                             ||||||||||||||||||||||||||||||||||||||||||
Db    431 TTACTCTAACTGGTCTCCATAGGTGTATTGCCTCAATAAGTCGGATGTTGAGTATACGCT 490

Qy   3225 ccacggtcactgtgggcctgtaacctgtctctttgtggatcgctggcaacctggcacagg 3284
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    491 CCACGGTCACTGTGGGCCTGTAACCTGTCTCTTTGTGGATCGCTGGCAACCTGGCACAGG 550

Qy   3285 ggggtctgggtcccaggacggcctgctctgcgtatgggatctgttcacgggagcctgcat 3344
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    551 GGGGTCTGGGTCCCAGGACGGCCTGCTCTGCGTATGGGATCTGTTCACGGGAGCCTGCAT 610

Qy   3345 gtataatatacaagctcacgacggagccgtcagctgcctggcctgtgcgcccagttacgt 3404
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    611 GTATAATATACAAGCTCACGACGGAGCCGTCAGCTGCCTGGCCTGTGCGCCCAGTTACGT 670

Qy   3405 aatctcgctaggcacggacgagaggatttgcgtatgggaacgatttcagggaaacctgtt 3464
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    671 AATCTCGCTAGGCACGGACGAGAGGATTTGCGTATGGGAACGATTTCAGGGAAACCTGTT 730

Qy   3465 gactaccatcaacatctcaaacgcatactcgagcctactgatgctaacaccgtcactatt 3524
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    731 GACTACCATCAACATCTCAAACGCATACTCGAGCCTACTGATGCTAACACCGTCACTATT 790

Qy   3525 ggttacgagcaaaatgggtaaggcctcattcttgattgccaatataagagggacagtaaa 3584
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    791 GGTTACGAGCAAAATGGGTAAGGCCTCATTCTTGATTGCCAATATAAGAGGGACAGTAAA 850

Qy   3585 taataaatttaattccaacacaggatctcttattgtgtgggatgtgcgcactgggcagcc 3644
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    851 TAATAAATTTAATTCCAACACAGGATCTCTTATTGTGTGGGATGTGCGCACTGGGCAGCC 910

Qy   3645 ggctcgcgaggtcaaactggactttgcaaacctgcagctctgtcccaaaataatgatgct 3704
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    911 GGCTCGCGAGGTCAAACTGGACTTTGCAAACCTGCAGCTCTGTCCCAAAATAATGATGCT 970

Qy   3705 tgcctgcgattcggtagtttgcgactacggaaatgagatccgcgtcgtccgctttcctat 3764
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    971 TGCCTGCGATTCGGTAGTTTGCGACTACGGAAATGAGATCCGCGTCGTCCGCTTTCCTAT 1030

Qy   3765 cgtggcagacaagtgccattaaagcgcaaaatttttaatttagcgtggttcgctagcacct 3824
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1031 CGTGGCAGACAAGTGCCATTAAAGCGCAAAATTTTAATTTAGCGTGGTTCGCTAGCACCT 1090

Qy   3825 aggaataagttgacttaaggctttaaaacgcctggaagtcattgacgcattcactatttt 3884
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1091 AGGAATAAGTTGACTTAAGGCTTTAAAACGCCTGGAAGTCATTGACGCATTCACTATTTT 1150

Qy   3885 atataaatatatacactattagggtccgcagcaacttacggttttaacacaagctgtacg 3944
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1151 ATATAAATATATACACTATTAGGGTCCGCAGCAACTTACGGTTTTAACACAAGCTGTACG 1210

Qy   3945 tatctcatctctagaattttgtgttagtttgtggacactaagtgtaacagctacgctccg 4004
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1211 TATCTCATCTCTAGAATTTTGTGTTAGTTTGTGGACACTAAGTGTAACAGCTACGCTCCG 1270

Qy   4005 gtaggttaaggaactaaactaaatgaatcagatatatacacatatattttcgcgtaatta 4064
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1271 GTAGGTTAAGGAACTAAACTAAATGAATCAGATATATACACATATATTTTCGCGTAATTA 1330

Qy   4065 tataaactacatagtgtcttaaagcgcctcagcctaatataaaatgactaaatgttaaaa 4124
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1331 TATAAACTACATAGTGTCTTAAAGCGCCTCAGCCTAATATAAAATGACTAAATGTTAAAA 1390
```

```
Qy   4125 taaa 4128
          ||||
Db   1391 TAAA 1394


RESULT   6
AC009255
LOCUS       AC009255   109699 bp    DNA             HTG       10-DEC-1999
DEFINITION  Drosophila melanogaster chromosome 2 clone BACR39E09 (D1043)
            RPCI-98 39.E.9 map 42A-42A strain y; cn bw sp, *** SEQUENCING IN
            PROGRESS ***, 103 unordered pieces.
ACCESSION   AC009255
VERSION     AC009255.3  GI:6554247
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 109699)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
            Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
            Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 109699)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Dec 10, 1999 this sequence version replaced gi:5734680.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 103 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1199: contig of 1199 bp in length
            *    1200     1279: gap of unknown length
            *    1280     1997: contig of 718 bp in length
            *    1998     2077: gap of unknown length
            *    2078     2686: contig of 609 bp in length
            *    2687     2766: gap of unknown length
            *    2767     3839: contig of 1073 bp in length
            *    3840     3919: gap of unknown length
            *    3920     4568: contig of 649 bp in length
            *    4569     4648: gap of unknown length
            *    4649     5320: contig of 672 bp in length
            *    5321     5400: gap of unknown length
            *    5401     6335: contig of 935 bp in length
            *    6336     6415: gap of unknown length
            *    6416     7390: contig of 975 bp in length
            *    7391     7470: gap of unknown length
            *    7471     8350: contig of 880 bp in length
            *    8351     8430: gap of unknown length
            *    8431     9510: contig of 1080 bp in length
            *    9511     9590: gap of unknown length
            *    9591    10436: contig of 846 bp in length
            *   10437    10516: gap of unknown length
            *   10517    11243: contig of 727 bp in length
            *   11244    11323: gap of unknown length
            *   11324    11946: contig of 623 bp in length
            *   11947    12026: gap of unknown length
            *   12027    12701: contig of 675 bp in length
            *   12702    12781: gap of unknown length
            *   12782    13062: contig of 281 bp in length
            *   13063    13142: gap of unknown length
            *   13143    14057: contig of 915 bp in length
            *   14058    14137: gap of unknown length
            *   14138    14805: contig of 668 bp in length
            *   14806    14885: gap of unknown length
            *   14886    15334: contig of 449 bp in length
            *   15335    15414: gap of unknown length
            *   15415    16203: contig of 789 bp in length
            *   16204    16283: gap of unknown length
            *   16284    17032: contig of 749 bp in length
            *   17033    17112: gap of unknown length
            *   17113    18667: contig of 1555 bp in length
            *   18668    18747: gap of unknown length
            *   18748    19961: contig of 1214 bp in length
            *   19962    20041: gap of unknown length
            *   20042    20778: contig of 737 bp in length
            *   20779    20858: gap of unknown length
            *   20859    21803: contig of 945 bp in length
            *   21804    21883: gap of unknown length
            *   21884    22739: contig of 856 bp in length
            *   22740    22819: gap of unknown length
            *   22820    23590: contig of 771 bp in length
            *   23591    23670: gap of unknown length
            *   23671    24575: contig of 905 bp in length
            *   24576    24655: gap of unknown length
            *   24656    25422: contig of 767 bp in length
            *   25423    25502: gap of unknown length
            *   25503    26176: contig of 674 bp in length
            *   26177    26256: gap of unknown length
            *   26257    26758: contig of 502 bp in length
            *   26759    26838: gap of unknown length
            *   26839    27932: contig of 1094 bp in length
            *   27933    28012: gap of unknown length
            *   28013    29308: contig of 1296 bp in length
            *   29309    29388: gap of unknown length
            *   29389    30496: contig of 1108 bp in length
            *   30497    30576: gap of unknown length
            *   30577    31728: contig of 1152 bp in length
            *   31729    31808: gap of unknown length
            *   31809    33080: contig of 1272 bp in length
            *   33081    33160: gap of unknown length
            *   33161    34333: contig of 1173 bp in length
            *   34334    34413: gap of unknown length
            *   34414    35478: contig of 1065 bp in length
            *   35479    35558: gap of unknown length
            *   35559    36465: contig of 907 bp in length
            *   36466    36545: gap of unknown length
            *   36546    37699: contig of 1154 bp in length
            *   37700    37779: gap of unknown length
            *   37780    38966: contig of 1187 bp in length
            *   38967    39046: gap of unknown length
            *   39047    40514: contig of 1468 bp in length
            *   40515    40594: gap of unknown length
            *   40595    41900: contig of 1306 bp in length
            *   41901    41980: gap of unknown length
            *   41981    43438: contig of 1458 bp in length
            *   43439    43518: gap of unknown length
            *   43519    44606: contig of 1088 bp in length
            *   44607    44686: gap of unknown length
            *   44687    46717: contig of 2031 bp in length
```

```
            *    46718    46797: gap of unknown length
            *    46798    48201: contig of 1404 bp in length
            *    48202    48281: gap of unknown length
            *    48282    49781: contig of 1500 bp in length
            *    49782    49861: gap of unknown length
            *    49862    51105: contig of 1244 bp in length
            *    51106    51185: gap of unknown length
            *    51186    52319: contig of 1134 bp in length
            *    52320    52399: gap of unknown length
            *    52400    54918: contig of 2519 bp in length
            *    54919    54998: gap of unknown length
            *    54999    56678: contig of 1680 bp in length
            *    56679    56758: gap of unknown length
            *    56759    59360: contig of 2602 bp in length
            *    59361    59440: gap of unknown length
            *    59441    62856: contig of 3416 bp in length
            *    62857    62936: gap of unknown length
            *    62937    66768: contig of 3832 bp in length
            *    66769    66848: gap of unknown length
            *    66849    70379: contig of 3531 bp in length
            *    70380    70459: gap of unknown length
            *    70460    78725: contig of 8266 bp in length
            *    78726    78805: gap of unknown length
            *    78806    79327: contig of 522 bp in length
            *    79328    79407: gap of unknown length
            *    79408    80062: contig of 655 bp in length
            *    80063    80142: gap of unknown length
            *    80143    80637: contig of 495 bp in length
            *    80638    80717: gap of unknown length
            *    80718    81225: contig of 508 bp in length
            *    81226    81305: gap of unknown length
            *    81306    81994: contig of 689 bp in length
            *    81995    82074: gap of unknown length
            *    82075    82836: contig of 762 bp in length
            *    82837    82916: gap of unknown length
            *    82917    83557: contig of 641 bp in length
            *    83558    83637: gap of unknown length
            *    83638    83920: contig of 283 bp in length
            *    83921    84000: gap of unknown length
            *    84001    84514: contig of 514 bp in length
            *    84515    84594: gap of unknown length
            *    84595    85404: contig of 810 bp in length
            *    85405    85484: gap of unknown length
            *    85485    86089: contig of 605 bp in length
            *    86090    86169: gap of unknown length
            *    86170    86848: contig of 679 bp in length
            *    86849    86928: gap of unknown length
            *    86929    87211: contig of 283 bp in length
            *    87212    87291: gap of unknown length
            *    87292    87990: contig of 699 bp in length
            *    87991    88070: gap of unknown length
            *    88071    88629: contig of 559 bp in length
            *    88630    88709: gap of unknown length
            *    88710    89373: contig of 664 bp in length
            *    89374    89453: gap of unknown length
            *    89454    89767: contig of 314 bp in length
            *    89768    89847: gap of unknown length
            *    89848    90407: contig of 560 bp in length
            *    90408    90487: gap of unknown length
            *    90488    90961: contig of 474 bp in length
            *    90962    91041: gap of unknown length

  Query Match             24.6%;  Score 1017.4;  DB 42;  Length 109699;
  Best Local Similarity   95.0%;  Pred. No. 7.3e-298;
  Matches 1084;  Conservative    0;  Mismatches    1;  Indels   56;  Gaps    1;

Qy   2982 aactcatgttcgaactggatccactggaagcctgggattaatgttgcagcagcagcgctg 3041
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30588 AGCTCATGTTCGAACTGGATCCACTGGAAGCCTGGGATTAATGTTGCAGCAGCAGCGCTG 30647

Qy   3042 tcagcaagaagcatcccagaagaccaccaaggaggaaatgaaaatcacattggagggtgt 3101
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30648 TCAGCAAGAAGCATCCCAGAAGACCACCAAGGAGGAAATGAAAATCACATTGGAGGGTGT 30707

Qy   3102 aagactagcccatcagcagccaatcacatgcatgcaggtcgttaacgacatggttttcac 3161
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30708 AAGACTAGCCCATCAGCAGCCAATCACATGCATGCAGGTCGTTAACGACATGGTTTTCAC 30767

Qy   3162 tggcagccaggatcacaccctca---------------------------------- 3184
          |||||||||||||||||||||||
Db  30768 TGGCAGCCAGGATCACACCCTCAAGGTAGGCATGAAACATTTTATTAAATCAACTTTCCT 30827

Qy   3185 ------------------aggtgtattgcctcaataagtcggatgttgagtatacgctc 3225
                            ||||||||||||||||||||||||||||||||||||||||||
Db  30828 TACTCTAACTGGTCTCCATAGGTGTATTGCCTCAATAAGTCGGATGTTGAGTATACGCTC 30887

Qy   3226 cacggtcactgtgggcctgtaacctgtctctttgtggatcgctggcaacctggcacaggg 3285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30888 CACGGTCACTGTGGGCCTGTAACCTGTCTCTTTGTGGATCGCTGGCAACCTGGCACAGGG 30947

Qy   3286 gggtctgggtcccaggacggcctgctctgcgtatgggatctgttcacgggagcctgcatg 3345
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30948 GGGTCTGGGTCCCAGGACGGCCTGCTCTGCGTATGGGATCTGTTCACGGGAGCCTGCATG 31007

Qy   3346 tataatatacaagctcacgacggagccgtcagctgcctggcctgtgcgcccagttacgta 3405
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31008 TATAATATACAAGCTCACGACGGAGCCGTCAGCTGCCTGGCCTGTGCGCCCAGTTACGTA 31067

Qy   3406 atctcgctaggcacggacgagaggatttgcgtatgggaacgatttcagggaaacctgttg 3465
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31068 ATCTCGCTAGGCACGGACGAGAGGATTTGCGTATGGGAACGATTTCAGGGAAACCTGTTG 31127

Qy   3466 actaccatcaacatctcaaacgcatactcgagcctactgatgctaacaccgtcactattg 3525
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31128 ACTACCATCAACATCTCAAACGCATACTCGAGCCTACTGATGCTAACACCGTCACTATTG 31187

Qy   3526 gttacgagcaaaatgggtaaggcctcattcttgattgccaatataagagggacagtaaat 3585
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31188 GTTACGAGCAAAATGGGTAAGGCCTCATTCTTGATTGCCAATATAAGAGGGACAGTAAAT 31247

Qy   3586 aataaatttaattccaacacaggatctcttattgtgtgggatgtgcgcactgggcagccg 3645
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31248 AATAAATTTAATTCCAACACAGGATCTCTTATTGTGTGGGATGTGCGCACTGGGCAGCCG 31307

Qy   3646 gctcgcgaggtcaaactggactttgcaaacctgcagctctgtcccaaaataatgatgctt 3705
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31308 GCTCGCGAGGTCAAACTGGACTTTGCAAACCTGCAGCTCTGTCCCAAAATAATGATGCTT 31367

Qy   3706 gcctgcgattcggtagtttgcgactacggaaatgagatccgcgtcgtccgctttcctatc 3765
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31368 GCCTGCGATTCGGTAGTTTGCGACTACGGAAATGAGATCCGCGTCGTCCGCTTTCCTATC 31427

Qy   3766 gtggcagacaagtgccattaaagcgcaaaatttaatttagcgtggttcgctagcaccta 3825
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31428 GTGGCAGACAAGTGCCATTAAAGCGCAAAATTTAATTTAGCGTGGTTCGCTAGCACCTA 31487

Qy   3826 ggaataagttgacttaaggctttaaaacgcctggaagtcattgacgcattcactatttta 3885
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31488 GGAATAAGTTGACTTAAGGCTTTAAAACGCCTGGAAGTCATTGACGCATTCACTATTTTA 31547

Qy   3886 tataaatatatacactattagggtccgcagcaacttacggttttaacacaagctgtacgt 3945
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31548 TATAAATATATACACTATTAGGGTCCGCAGCAACTTACGGTTTTAACACAAGCTGTACGT 31607

Qy   3946 atctcatctctagaattttgtgttagtttgtggacactaagtgtaacagctacgctccgg 4005
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31608 ATCTCATCTCTAGAATTTTGTGTTAGTTTGTGGACACTAAGTGTAACAGCTACGCTCCGG 31667

Qy   4006 taggttaaggaactaaactaaatgaatcagatatatacacatatattttcgcgtaattat 4065
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31668 TAGGTTAAGGAACTAAACTAAATGAATCAGATATATACACATATATTTTCGCGTAATTAT 31727

Qy   4066 a 4066
          |
Db  31728 A 31728
```

```
RESULT   7
CNS01P6Y
LOCUS       CNS01P6Y      790 bp    DNA             STS       18-FEB-2000
DEFINITION  Anopheles gambiae STS SP6 end of clone 27019 of NotreDame1 library
            from strain PEST of Anopheles gambiae (African malaria mosquito),
            sequence tagged site.
ACCESSION   AL154442
VERSION     AL154442.1  GI:7015361
KEYWORDS    STS.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
            Culicoidea; Culicidae; Anopheles.
REFERENCE   1  (bases 1 to 790)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
REFERENCE   2  (bases 1 to 790)
  AUTHORS   Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
            Roux, Paris 75015, France
COMMENT     This clone is from an A. gambiae BAC library provided by F.H.
            Collins and sequenced by Genoscope in collaboration with the
            Laboratory of Biochem. and Biol. Molec. of Insects, Institut
            Pasteur.
FEATURES             Location/Qualifiers
     source          1. .790
                     /organism="Anopheles gambiae"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="27019"
                     /clone_lib="NotreDame1"
                     /note="end : SP6"
BASE COUNT      159 a    194 c    247 g    178 t     12 others
ORIGIN


  Query Match             6.4%;  Score 263.2;  DB 13;  Length 790;
  Best Local Similarity   64.4%;  Pred. No. 1.9e-68;
  Matches  382;  Conservative    6;  Mismatches  205;  Indels    0;  Gaps    0;

Qy    923 cgtcggctgaagagccgacgaccataacttacatcttttatccaggagagtacaggatgt 982
          || |  |  |  ||| |    | || |  |||||||| || || || || | ||    |
Db      6 CGGCCCCGAAGCAGCAGCRATCGATCATGTACATCTTCTACCCGGGCGAATTCAACTGGA 65

Qy    983 gggagctggtgccttacacagtggcctttatgttggtgtttgcttatgtgtacttctctg 1042
          |||| || ||||  |     | || |||    || ||||   ||| || ||||| || |
Db     66 AGGAGGTGCTGCCCCAGTGTATCGCGTTTTGTGTGCTGTTCATTTACGTCTACTTTTCCG 125

Qy   1043 ttcgaaaaatcgatgtatttcgttcccgctttttgctggccttatgtagcgtaatcacca 1102
          | || |||||||| |   | || || ||| |: |||| || |  |||  |||  | |||
Db    126 TCCGCAAAATCGAAGCCATCCGGTCGCGCATYATGCTAGCGTCCTGTGCCGTGCTGACCG 185

Qy   1103 cagccgggagcttggccatgtcccttggcttgtgtttcttctttggcctgacaatttcgc 1162
               || ||||||   ||||: || ||  |||||||||||||| || || || ||:
Db    186 TGCTGGGCAGCTTGCTGATGTYACTGGGGCTGTGTTTCTTCTTCGGGCTTACGATYAGCT 245

Qy   1163 tgcagtcaaaggacattttcccctaccttgtaatccttgtgggattggaaaatagcttgg 1222
          | ||||| ||||   | |  || ||||| || || || |||||  |||| ||     | |
Db    246 TCCAGTCGAAGGGTGTCTATCCGTACCTGGTCATACTGGTGGGGCTGGAGAACGTGCTCG 305

Qy   1223 tgatcacaaagagcgtagtctcaatggacgagacattcgacgtgaagatccgcgtggcgc 1282
          | || || ||||| ||  || | | |||| | |  :| ||||| ||||| || || || |
Db    306 TTATTACGAAGAGTGTGCTCACGACGGACAACAATYTGGACGTWAAGATACGGGTCGCAC 365

Qy   1283 aggctcttagcaaggagggttggcatatatccaagactcttttgacggagataacaattt 1342
          | | ||| |||  |||||| |||   ||    ||||   |  |||||| || || ||
Db    366 AAGGTCTGAGCCGGGAGGGCTGGTCCATTATGAAGAACTTGCTGACGGAAATTACCATCC 425

Qy   1343 tgacaattggtcttgctactttcgtgcccgtcatccaggagttttgtatctttgccatag 1402
          | ||    || ||||| || |||||:|| || || ||||||||||:||||||||  || |
Db    426 TTACCGGGGGGCTTGCGACGTTCGTKCCGGTAATACAGGAGTTTTKTATCTTTGNGATGG 485

Qy   1403 tcggcttgctttccgattttatgctacagatgctgctcttctcaacaatactggccatga 1462
          | ||| |  | || || ||| |||| |||||||| ||||| | ||  | |||||  ||
Db    486 TGGGCCTCGTGTCGGACTTTCTGCTGCAGATGCTGTTCTTCGCCACCGTGCTGGCGCTGG 545

Qy   1463 acattaagcggaccgagtatacggcggaggccaagcaccttcctaagatgttg 1515
          ||||||||||    ||||| |  |||||||    ||  || || |||||| ||
Db    546 ACATTAAGCGAGTTGAGTACAGTGCGGAGGTGCGGCGACTGCCCAAGATGCTG 598


RESULT   8
D83782
LOCUS       D83782       3976 bp    mRNA            PRI       10-JUL-1997
DEFINITION  Human mRNA for KIAA0199 gene, partial cds.
ACCESSION   D83782
VERSION     D83782.1  GI:1228046
KEYWORDS    KIAA0199.
SOURCE      Homo sapiens male myeloblast cell_line:KG-1 cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3976)
  AUTHORS   Nomura,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-1996) to the DDBJ/EMBL/GenBank databases. Nobuo
            Nomura, Kazusa DNA Research Institute, Gene Structure 1; 1532-3
            Yana, Kisarazu, Chiba 292, Japan (E-mail:cdnainfo@kazusa.or.jp,
            URL:http://www.kazusa.or.jp, Tel:0438-52-3930, Fax:0438-52-3931)
REFERENCE   2  (bases 1 to 3976)
  AUTHORS   Nagase,T., Seki,N., Ishikawa,K., Tanaka,A. and Nomura,N.
  TITLE     Prediction of the coding sequences of unidentified human genes. V.
            The coding sequences of 40 new genes (KIAA 0161 - KIAA 0200)
            deduced by analysis of cDNA clones from human cell line KG-1
  JOURNAL   DNA Res. (1995) In press
REFERENCE   3  (sites)
  AUTHORS   Nagase,T., Seki,N., Ishikawa,K., Tanaka,A. and Nomura,N.
  TITLE     Prediction of the coding sequences of unidentified human genes. V.
            The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
            analysis of cDNA clones from human cell line KG-1
  JOURNAL   DNA Res. 3 (1), 17-24 (1996)
  MEDLINE   96281124
REFERENCE   4  (sites)
  AUTHORS   Hua,X., Nohturfft,A., Goldstein,J.L. and Brown,M.S.
  TITLE     Sterol resistance in CHO cells traced to point mutation in SREBP
            cleavage-activating protein
  JOURNAL   Cell 87 (3), 415-426 (1996)
  MEDLINE   97053781
FEATURES             Location/Qualifiers
     source          1. .3976
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="KG-1"
                     /cell_type="myeloblast"
                     /chromosome="3"
                     /sex="male"
     gene            1. .3835
                     /gene="KIAA0199"
     CDS             <1. .3835
                     /gene="KIAA0199"
                     /note="the KIAA0199 gene is expressed ubiquitously.; the
                     KIAA0199 protein shows similarity to sea urchin
                     hydroxymethylglutalyl-CoA reductase, and retains 8
                     hydrophobic domains."
                     /codon_start=2
                     /protein_id="BAA12111.1"
                     /db_xref="GI:1228047"
                     /translation="LTERLREKISRAFYNHGLLCASYPIPIILFTGFCILACCYPLLK
                     LPLPGTGPVEFTTPVKDYSPPPVDSDRKQGEPTEQPEWYVGAPVAYVQQIFVKSSVFP
                     WHKNLLAVDVFRSPLSRAFQLVEEIRNHVLRDSSGIRSLEELCLQVTDLLPGLRKLRN
```

```
                     LLPEHGCLLLSPGNFWQNDWERFHADPDIIGTIHQHEPKTLQTSATLKDLLFGVPGKY
                     SGVSLYTRKRMVSYTITLVFQHYHAKFLGSLRARLMLLHPSPNCSLRAESLVHVHFKE
                     EIGVAELIPLVTTYIILFAYIYFSTRKIDMVKSKWGLALAAVVTVLSSLLMSVGLCTL
                     FGLTPTLNGGEIFPYLVVVIGLENVLVLTKSVVSTPVDLEVKLRIAQGLSSESWSIMK
                     NMATELGIILIGYFTLVPAIQEFCLFAVVGLVSDFFLQMLFFTTVLSIDIRRMELADL
                     NKRLPPEACLPSAKPVGQPTRYERQLAVRPSTPHTITLQPSSFRNLRLPKRLRVVYFL
                     ARTRLAQRLIMAGTVVWIGILVYTDPAGLRNYLAAQVTEQSPLGEGALAPMPVPSGML
                     PPSHPDPAFSIFPPDAPKLPENQTSPGESPERGGPAEVVHDSPVPEVTWGPEDEELWR
                     KLSFRHWPTLFSYYNITLAKRYISLLPVIPVTLRLNPREALEGRHPQDGRSAWPPPGP
                     IPAGHWEAGPKGPGGVQAHGDVTLYKVAALGLATGIVLVLLLLCLYRVLCPRNYGQLG
                     GGPGRRRRGELPCDDYGYAPPETEIVPLVLRGHLMDIECLASDGMLLVSCCLAGHVCV
                     WDAQTGDCLTRIPRPGRQRRDSGVGSGLEAQESWERLSDGGKAGPEEPGDSPPLRHRP
                     RGPPPPSLFGDQPDLTCLIDTNFSAQPRSSQPTQPEPRHRAVCGRSRDSPGYDFSCLV
                     QRVYQEEGLAAVCTPALRPPSPGPVLSQAPEDEGGSPEKGSPSLAWAPSAEGSIWSLE
                     LQGNLIVVGRSSGRLEVWDAIEGVLCCSSEEVSSGITALVFLDKRIVAARLNGSLDFF
                     SLETHTALSPLQFRGTPGRGSSPASPVYSSSDTVACHLTHTVPCAHQKPITALKAAAG
                     RLVTGSQDHTLRVFRLEDSCCLFTLQGHSGAITTVYIDQTMVLASGGQDGAICLWDVL
                     TGSRVSHVFAHRGDVTSLTCTTSCVISSGLDDLISIWDRSTGIKFYSIQQDLGCGASL
                     GVISDNLLVTGGQGCVSFWDLNYGDLLQTVYLGKNSEAQPARQILVLDNAAIVCNFGS
                     ELSLVYVPSVLEKLD"
    3'UTR            3836. .3976
BASE COUNT      725 a   1269 c   1172 g    810 t
ORIGIN
```

```
  Query Match             4.7%;  Score 193.6;  DB 9;  Length 3976;
  Best Local Similarity   51.3%;  Pred. No. 4.7e-47;
  Matches  597;  Conservative    0;  Mismatches  534;  Indels   33;  Gaps    5;

Qy    351 gcctccccttctgccctgggcgcagagcagccccgccttttttctacgtccagcagattac 410
          |||| |    | |||   |  | |     |  || |  |    || ||||||||||||
Db    226 GCCTACTGAGCAGCCTGAGTGGTATGTGGGTGCCCCGGTGGCTTATGTCCAGCAGATATT 285

Qy    411 tctgcgaaccagtgttctcccgtggacggagggaatgcagcttatggatgcgtttcgtgc 470
          | ||    ||   ||  | || |||    ||    | | |    | ||||  ||||||| |
Db    286 TGTGAAGTCCTCAGTGTTTCCCTGGCACAAGAACCTCCTGGCAGTAGATGTATTTCGTTC 345

Qy    471 gccgctacacgaagtttttaaattgcttgaaattgtgcgcaatcaccagagcagcgaaaa 530
          ||  |   |   |  ||  || || | ||     | || || |||  |   || || |
Db    346 ACCTTTGTCCCGGGCATTCCAACTGGTGGAGGAGATCCGGAACCACGTGCTGAGAGACAG 405

Qy    531 c------aaacgtaccctggagcacaactgcctacatgtagacaacgtaaagcgcggaac 584
          |      |   | | | ||||| |    || || || ||   | || |   ||  ||
Db    406 CTCTGGGATCAGGAGCTTGGAGGAGTTGTGTCTGCAAGTGACCGACCTGCTGCCAGG--- 462

Qy    585 acacgggcagctggaccagatctttccggagtatggctgcctgctgctctcgcccgccaa 644
          |   || ||||     |  |  | || ||| |||| ||||||||||| || || |  ||
Db    463 CCTTAGGAAGCTCAGGAACCTACTCCCTGAGCATGGATGCCTGCTGCTGTCCCCTGGGAA 522

Qy    645 cctgtggacgcagaactctcagaactttactcgggacacaaacatcctgaacacgatatt 704
          | | |||  | |  |||   |  |||   |  ||  |  ||||| |    || ||
Db    523 CTTCTGGCAGAATGACTGGGAACGCTTCCATGCTGATCCTGACATCATTGGGACCATCCA 582

Qy    705 tcagtac---cataacctacagaaatcaaaagtttccgcggcggaaatgctgtttggatt 761
          |||  ||   | |||    | | |  |   ||   |       ||  || | |||||  |
Db    583 CCAGCACGAGCCTAAAACCCTGCAGACTTCAGCCACACTCAAAGACTTGTTATTTGGTGT 642

Qy    762 acccatgcaggacactggattcaagcgctatccattgcgcgctcggtcgcgtattataca 821
          ||   | || |||  ||  | |  | |||          | |  ||  | | ||  |
Db    643 TCCTGGGAAGTACAGCGGGGTGAGCCTCTA---------CACCAGGAAGAGGATGGTCTC 693

Qy    822 gtatgccttgacgttattcctcaagcacaacgatatggagtatctggacactctaaagga 881
          ||  || | ||   |  || || ||||| || ||    |||  |||| ||  ||    |
Db    694 CTACACCATCACCCTGGTCTTCCAGCACTACCATGCCAAGTTCCTGGGCAGCCTGCGTGC 753

Qy    882 aaagctgctgcgacactatcccccactcccgttggctagtgcgtcggctgaagagccgac 941
              ||| |||  |   | |||   | |   |       |  | |||
Db    754 CCGCCTGATGCTTCTGCACCCCAGCCCCAACTGCAGCCTTCGGGCGG----------A 801

Qy    942 gaccataacttacatcttttatccaggagagtacaggatgtgggagctggtgccttacac 1001
          || | |     || |    |     |  |||    |  |    |||||  | ||
Db    802 GAGCCTGGTCCACGTGCACTTCAAGGAGGAGATTGGTGTCGCTGAGCTCATCCCCCTTGT 861
```

```
Qy   1002 agtggcctttatgttggtgtttgcttatgtgtacttctctgttcgaaaaatcgatgtatt 1061
              |||  ||  |  |||||| ||  | ||||||||    || || |||||  |  |
Db    862 GACCACCTACATCATCTTGTTTGCCTACATCTACTTCTCCACGCGGAAGATCGACATGGT 921

Qy   1062 tcgttcccgctttttgctggccttatgtagcgtaatcaccacagccgggagcttggccat 1121
            |||  |    |||||||  |    |  |||  ||||      | |      ||  |||
Db    922 CAAGTCCAAGTGGGGCTGGCCCTGGCTGCCGTGGTCACAGTGCTCAGCTCGCTGCTCAT 981

Qy   1122 gtcccttggcttgtgtttcttctttggcctgacaatttcgctgcagtcaaaggacatttt 1181
          |||  | ||  | ||     |||| ||||||||     | ||  |       || |||||
Db    982 GTCTGTGGGACTCTGCACACTCTTCGGCCTGACGCCCACCCTCAATGGCGGCGAGATTTT 1041

Qy   1182 cccctaccttgtaatccttgtgggattggaaaatagcttggtgatcacaaagagcgtagt 1241
          ||||||||||||  |  || | || || || |||   |||||| |||| |||   || ||
Db   1042 CCCCTACCTTGTGGTGGTTATTGGGTTAGAGAATGTGTTGGTGCTCACCAAGTCTGTGGT 1101

Qy   1242 ctcaatggacgagacattcgacgtgaagatccgcgtggcgcaggctcttagcaaggaggg 1301
          |||||     |      | || |||||| | ||  | || || |  || |||  ||| |
Db   1102 CTCAACCCCGGTAGACCTGGAGGTGAAGCTGCGGATCGCCCAAGGCCTAAGCAGCGAGAG 1161

Qy   1302 ttggcatatatccaagactcttttgacggagataacaattttgacaattggtcttgctac 1361
           |||   ||    ||||   |    |||||| |    ||  |    || ||       ||
Db   1162 CTGGTCCATCATGAAGAACATGGCCACGGAGCTGGGCATCATCCTCATCGGCTACTTCAC 1221

Qy   1362 tttcgtgcccgtcatccaggagttttgtatctttgccatagtcggcttgctttccgattt 1421
            | |||||| |||||||||||| ||| |||||||  | || ||  || | || || ||
Db   1222 CCTAGTGCCCGCCATCCAGGAGTTCTGTCTCTTTGCTGTCGTGGGGCTGGTGTCTGACTT 1281

Qy   1422 tatgctacagatgctgctcttctcaacaatactggccatgaacattaagcggaccgagta 1481
          | || |||||||||| | ||| | ||  | |||  |||| ||||||   ||||  |||
Db   1282 CTTCCTTCAGATGCTGTTTTTCACCACTGTCCTGTCCATTGACATTCGCCGGATGGAGCT 1341

Qy   1482 tacggcggaggccaagcaccttcc 1505
          | |    |  |||||  || ||
Db   1342 AGCAGACCTGAACAAGCGACTGCC 1365
```

```
RESULT   9
CGU67060
LOCUS       CGU67060     4119 bp    mRNA            ROD       19-NOV-1996
DEFINITION  Cricetulus griseus SREBP cleavage activating protein (SCAP) mRNA,
            complete cds.
ACCESSION   U67060
VERSION     U67060.1  GI:1675219
KEYWORDS    .
SOURCE      Chinese hamster.
  ORGANISM  Cricetulus griseus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
REFERENCE   1  (bases 1 to 4119)
  AUTHORS   Hua,X., Nohturfft,A., Goldstein,J.L. and Brown,M.S.
  TITLE     Sterol resistance in CHO cells traced to point mutation in SREBP
            cleavage-activating protein
  JOURNAL   Cell 87 (3), 415-426 (1996)
  MEDLINE   97053781
REFERENCE   2  (bases 1 to 4119)
  AUTHORS   Hua,X., Nohturfft,A., Goldstein,J.L. and Brown,M.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-AUG-1996) Molecular Genetics, The University of Texas
            Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX
            75235, USA
FEATURES             Location/Qualifiers
     source          1. .4119
                     /organism="Cricetulus griseus"
                     /db_xref="taxon:10029"
                     /tissue_type="ovary"
     gene            120. .3950
                     /gene="SCAP"
     CDS             120. .3950
                     /gene="SCAP"
                     /note="Homologous to the transmembrane domains of HMG CoA
                     reductase, containing WD repeats"
                     /codon_start=1
```

```
                     /product="SREBP cleavage activating protein"
                     /protein_id="AAB19103.1"
                     /db_xref="GI:1675220"
                     /translation="MTLTERLREKISQAFYNHGLLCASYPIPIILFTGLCILACCYPL
                     LKLPLPGTGPVEFSTPVKDYSPPPVDSDHKQGEPSEQPEWYVGAPVAYIQQIFVKSSV
                     SPWHKNLLAVDVFRLPLSRAFQLVEEIRNHVLRDSSGTKSLEEVCLQVTDLLPGLRKL
                     RNLLPEHGCLLLSPGNFWQNDWERFHADPDIIGTIHQHEPKTLQTSATLKDLLFGVPG
                     KYSGVSLYTRKRTVSYTITLVFQRYHAKFLSSLRARLMLLHPSPNCSLRAENLVHVHF
                     KEEIGIAELIPLVTTYIILFAYIYFSTRKIDMVKSKWGLALAAVVTVLSSLLMSVGLC
                     TLFGLTPTLNGGEIFPYLVVVIGLENVLVLTKSVVSTPVDLEVKLRIAQGLSSESWSI
                     MKNVATELGIILIGYFTLVPAIQEFCLFAVVGLVSDFFLQMFFFTTVLSIDIRRMELA
                     DLNKRLPPESCLPSAKPVGRPARYERQLAVRPAMPHTITLQPSSFRNLRLPKRLRVIY
                     FLARTRLAQRLIMAGTVVWIGILVYTDPAGLRTYLAAQVTEQSPLGEGSLGPMPVPSG
                     VLPASRPDPAFSIFPPDAPKLPENQTVPGELPEHAAPAEGVHDSRAPEVTWGPEDEEL
                     WRRLSFRHWPTLFNYYNITLAKRYISLLPVIPVTLRLNPQEALEGRQPQDGRSAWAPP
                     ESLPAGLWEAGPKGPGGTQAHGDITLYKVAALGLAAGIVLVLLLLCLYRVLCPRNYGQ
                     PGGGAGRRRRGELPCDDYGYAPPETEIVPLVLRGHLMDIECLASDGMLLVSCCLAGQV
                     CVWDAQTGDCLTRIPRPGSRRDSCGGGAFETQENWERLSDGGKTSPEEPGESPPLRHR
                     PRGPPQPALFGDQPDLTCLIDTNFSVQLPPEPTQPEPRHRAGCGRARDSGYDFSRLVQ
                     RVYQEEGLAAVRMPALRPPSPGSPLPQASQEDGAAPEKGSPPLAWAPSTAGSIWSLEL
                     QGNLIVVGRSSGRLEVWDAIEGVLCCSNDEVSSGITALVFLDRRIVAARLNGSLDFFS
                     LETHTSLSPLQFRGTPGRGSSPSSSVYSSSNTVACHLTHTVPCAHQKPITALRAAAGR
                     LVTGSQDHTLRVFRLEDSCCLFTLQGHSGAITTVYIDQTMVLASGGQDGAICLWDVLT
                     GSRVSHTFAHRGDVTSLTCTTSCVISSGLDDLINIWDRSTGIKLYSIQQDLGCGASLG
                     VISDNLLVTGGQGCVSFWDLNYGDLLQTVYLGKNSEAQPARQILVLDNAAIVCNFGSE
                     LSLVYVPSVLEKLD"
BASE COUNT      824 a   1255 c   1153 g    887 t
ORIGIN


  Query Match             4.5%;  Score 187.4;  DB 12;  Length 4119;
  Best Local Similarity   51.5%;  Pred. No. 3.7e-45;
  Matches  577;  Conservative    0;  Mismatches  511;  Indels   33;  Gaps    5;

Qy     394 tacgtccagcagattactctgcgaaccagtgttctcccgtggacggagggaatgcagctt 453
           ||| |||| |||||   | ||   |    ||    || |||    |     | | |
Db     393 TACATCCAACAGATATTTGTGAAGTCATCGGTATCTCCCTGGCACAAAAATCTTCTGGCA 452

Qy     454 atggatgcgtttcgtgcgccgctacacgaagtttttaaattgcttgaaattgtgcgcaat 513
           | |||| ||| ||   ||| ||   |   |  ||  || || | ||     | || |||
Db     453 GTAGATGTGTTCCGATTGCCTCTGTCCCGGGCATTCCAACTGGTGGAGGAAATCCGGAAT 512

Qy     514 caccagagcagcgaaaacaaacgtac------cctggagcacaactgcctacatgtagac 567
           |||      || || | |  | | ||      |||||| |    ||||| || ||   |
Db     513 CACGTACTGAGAGACAGCTCAGGGACCAAGAGCCTGGAGGAAGTTTGCCTTCAGGTGACC 572

Qy     568 aacgtaaagcgcggaacacacgggcagctggaccagatctttccggagtatggctgcctg 627
           || |   ||  ||   || | || | ||     |  |  |||| ||  ||||||||||
Db     573 GACTTGCTGCCAGG---ACTCAGGAAACTCCGAAACCTACTTCCTGAACATGGCTGCCTG 629

Qy     628 ctgctctcgcccgccaacctgtggacgcagaactctcagaactttactcgggacacaaac 687
           ||||| || || |  ||| | |||  | |  |||   ||   ||   |  ||| |  ||
Db     630 CTGCTGTCCCCTGGAAACTTCTGGCAGAATGACTGGGAGCGATTCCATGCTGACCCTGAC 689

Qy     688 atcctgaacacgatatttcag---taccataacctacagaaatcaaaagtttccgcggcg 744
           ||| |    || ||   ||||    | |  ||    | | |  ||  ||   |
Db     690 ATCATTGGGACTATCCATCAGCATGAGCCCAAAACTCTGCAGACATCAGCCACACTCAAA 749

Qy     745 gaaatgctgtttggattacccatgcaggacactggattcaagcgctatccattgcgcgct 804
           ||  ||||||||||  | ||   | || ||| |||  | |  | |||         | |
Db     750 GACCTGCTGTTTGGTGTGCCTGGGAAGTACAGTGGGGTGAGCCTCTA---------CACA 800

Qy     805 cggtcgcgtattatacagtatgccttgacgttattcctcaagcacaacgatatggagtat 864
           ||  | | |   |   |||  || | ||  |  || || ||| | || ||    ||| |
Db     801 AGGAAGAGGACGGTCTCGTACACCATCACCCTGGTCTTCCAGCGCTACCATGCCAAGTTT 860

Qy     865 ctggacactctaaaggaaaagctgctgcgacactatcccccactcccgttggctagtgcg 924
           |||  ||  |||    |   ||| |||  |   | ||     | |  |       | |
Db     861 CTGAGCAGCCTACGTGCCCGGCTCATGCTCCTGCACCCTAGCCCCAACTGCAGCCTTCGG 920

Qy     925 tcggctgaagagccgacgaccataacttacatcttttatccaggagagtacaggatgtgg 984
           | |             || | |    || |   |    ||  |||    | ||   |
Db     921 GCCG------------AGAACCTGGTCCACGTGCACTTCAAAGAGGAGATTGGCATTGCG 968

Qy     985 gagctggtgccttacacagtggccttatgttggtgtttgcttatgtgtacttctctgtt 1044
           |||||  | ||   |       |||  ||  |  ||||||| ||  | ||||||||
Db     969 GAGCTCATCCCCCTCGTGACCACCTACATTATCCTGTTTGCCTACATCTACTTCTCCACA 1028

Qy    1045 cgaaaaatcgatgtatttcgttcccgctttttgctggccttatgtagcgtaatcaccaca 1104
           || || |||||  |  |    ||    |     || ||| |       ||  ||||
Db    1029 CGTAAGATCGACATGGTCAAGTCTAAGTGGGGACTCGCCCTGGCAGCTGTGGTCACAGTG 1088

Qy    1105 gccgggagcttggccatgtcccttggcttgtgtttcttctttggcctgacaatttcgctg 1164
              |     ||  ||||||  | ||  | ||     ||||||||||||||    | ||
Db    1089 CTTAGCTCTCTGCTCATGTCTGTGGGGCTCTGCACTCTCTTTGGCCTGACACCCACACTC 1148

Qy    1165 cagtcaaaggacattttcccctaccttgtaatccttgtgggattggaaaatagcttggtg 1224
              |       || |||||||| ||||| ||  |  || | || || || |||   ||||||
Db    1149 AATGGCGGTGAGATTTTCCCATACCTGGTGGTGGTTATTGGGTTAGAGAATGTGTTGGTG 1208

Qy    1225 atcacaaagagcgtagtctcaatggacgagacattcgacgtgaagatccgcgtggcgcag 1284
            ||||  |||   || || ||||     | |    |||| |||||| | ||  | || ||
Db    1209 CTCACCAAGTCGGTTGTATCAACCCCAGTGGACCTCGAGGTGAAGCTTCGGATTGCACAA 1268

Qy    1285 gctcttagcaaggagggttggcatatatccaagactcttttgacggagataacaattttg 1344
           |   | ||||  ||| | |||   ||    ||||   |    || ||  |    ||  |
Db    1269 GGCTTAAGCAGTGAGAGCTGGTCCATCATGAAGAACGTGGCAACTGAACTGGGCATCATC 1328

Qy    1345 acaattggtcttgctactttcgtgcccgtcatccaggagttttgtatctttgccatagtc 1404
              |||||   |   ||  | ||||| |  |||||||||| ||  |||||||  | ||
Db    1329 CTCATTGGCTATTTCACCCTTGTGCCTGCTATCCAGGAGTTCTGCCTCTTTGCTGTTGTG 1388

Qy    1405 ggcttgctttccgattttatgctacagatgctgctcttctcaacaatactggccatgaac 1464
           ||| || | || || ||  | || |||||| |  | ||| | ||  | ||| | ||  ||
Db    1389 GGCCTGGTGTCTGACTTCTTCCTCCAGATGTTCTTTTTCACCACTGTCCTGTCGATCGAC 1448

Qy    1465 attaagcggaccgagtatacggcggaggccaagcaccttcc 1505
           |||  |||  |  |||    |||    |  |||||  || ||
Db    1449 ATTCGGCGAATGGAGCTAGCGGACCTGAACAAGCGTCTGCC 1489


RESULT  10
AC008339
LOCUS       AC008339   118474 bp    DNA             HTG       06-AUG-1999
DEFINITION  Drosophila melanogaster chromosome 2 clone BACR13P06 (D917) RPCI-98
            13.P.6 map 42A-42A strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***, 97 unordered pieces.
ACCESSION   AC008339
VERSION     AC008339.1  GI:5670396
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 118474)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 118474)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
```

Rubin,G.M.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://www.fruitfly.org/sequence/) or send email to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 97 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	1154: contig of 1154 bp in length
*	1155	1234: gap of unknown length
*	1235	1978: contig of 744 bp in length
*	1979	2058: gap of unknown length
*	2059	2763: contig of 705 bp in length
*	2764	2843: gap of unknown length
*	2844	3474: contig of 631 bp in length
*	3475	3554: gap of unknown length
*	3555	4790: contig of 1236 bp in length
*	4791	4870: gap of unknown length
*	4871	5691: contig of 821 bp in length
*	5692	5771: gap of unknown length
*	5772	7104: contig of 1333 bp in length
*	7105	7184: gap of unknown length
*	7185	8019: contig of 835 bp in length
*	8020	8099: gap of unknown length
*	8100	8783: contig of 684 bp in length
*	8784	8863: gap of unknown length
*	8864	9614: contig of 751 bp in length
*	9615	9694: gap of unknown length
*	9695	10352: contig of 658 bp in length
*	10353	10432: gap of unknown length
*	10433	11036: contig of 604 bp in length
*	11037	11116: gap of unknown length
*	11117	11859: contig of 743 bp in length
*	11860	11939: gap of unknown length
*	11940	12630: contig of 691 bp in length
*	12631	12710: gap of unknown length
*	12711	13410: contig of 700 bp in length
*	13411	13490: gap of unknown length
*	13491	14724: contig of 1234 bp in length
*	14725	14804: gap of unknown length
*	14805	15537: contig of 733 bp in length
*	15538	15617: gap of unknown length
*	15618	16875: contig of 1258 bp in length
*	16876	16955: gap of unknown length
*	16956	18332: contig of 1377 bp in length
*	18333	18412: gap of unknown length
*	18413	19139: contig of 727 bp in length
*	19140	19219: gap of unknown length
*	19220	20179: contig of 960 bp in length
*	20180	20259: gap of unknown length
*	20260	21313: contig of 1054 bp in length
*	21314	21393: gap of unknown length
*	21394	22510: contig of 1117 bp in length
*	22511	22590: gap of unknown length
*	22591	23469: contig of 879 bp in length
*	23470	23549: gap of unknown length
*	23550	24814: contig of 1265 bp in length
*	24815	24894: gap of unknown length
*	24895	26082: contig of 1188 bp in length
*	26083	26162: gap of unknown length
*	26163	27865: contig of 1703 bp in length
*	27866	27945: gap of unknown length
*	27946	29037: contig of 1092 bp in length
*	29038	29117: gap of unknown length
*	29118	30403: contig of 1286 bp in length
*	30404	30483: gap of unknown length
*	30484	31874: contig of 1391 bp in length
*	31875	31954: gap of unknown length
*	31955	33181: contig of 1227 bp in length
*	33182	33261: gap of unknown length
*	33262	34653: contig of 1392 bp in length
*	34654	34733: gap of unknown length
*	34734	36014: contig of 1281 bp in length
*	36015	36094: gap of unknown length
*	36095	37591: contig of 1497 bp in length
*	37592	37671: gap of unknown length
*	37672	38920: contig of 1249 bp in length
*	38921	39000: gap of unknown length
*	39001	40308: contig of 1308 bp in length
*	40309	40388: gap of unknown length
*	40389	41517: contig of 1129 bp in length
*	41518	41597: gap of unknown length
*	41598	43130: contig of 1533 bp in length
*	43131	43210: gap of unknown length
*	43211	44533: contig of 1323 bp in length
*	44534	44613: gap of unknown length
*	44614	45686: contig of 1073 bp in length
*	45687	45766: gap of unknown length
*	45767	47157: contig of 1391 bp in length
*	47158	47237: gap of unknown length
*	47238	48766: contig of 1529 bp in length
*	48767	48846: gap of unknown length
*	48847	50558: contig of 1712 bp in length
*	50559	50638: gap of unknown length
*	50639	52587: contig of 1949 bp in length
*	52588	52667: gap of unknown length
*	52668	54127: contig of 1460 bp in length
*	54128	54207: gap of unknown length
*	54208	56171: contig of 1964 bp in length
*	56172	56251: gap of unknown length
*	56252	57717: contig of 1466 bp in length
*	57718	57797: gap of unknown length
*	57798	59216: contig of 1419 bp in length
*	59217	59296: gap of unknown length
*	59297	61485: contig of 2189 bp in length
*	61486	61565: gap of unknown length
*	61566	63346: contig of 1781 bp in length
*	63347	63426: gap of unknown length
*	63427	65078: contig of 1652 bp in length
*	65079	65158: gap of unknown length
*	65159	67680: contig of 2522 bp in length
*	67681	67760: gap of unknown length
*	67761	69367: contig of 1607 bp in length
*	69368	69447: gap of unknown length
*	69448	72886: contig of 3439 bp in length
*	72887	72966: gap of unknown length
*	72967	74413: contig of 1447 bp in length
*	74414	74493: gap of unknown length
*	74494	76339: contig of 1846 bp in length
*	76340	76419: gap of unknown length
*	76420	80413: contig of 3994 bp in length
*	80414	80493: gap of unknown length
*	80494	88132: contig of 7639 bp in length
*	88133	88212: gap of unknown length
*	88213	88883: contig of 671 bp in length
*	88884	88963: gap of unknown length
*	88964	89611: contig of 648 bp in length
*	89612	89691: gap of unknown length
*	89692	90372: contig of 681 bp in length
*	90373	90452: gap of unknown length
*	90453	91133: contig of 681 bp in length
*	91134	91213: gap of unknown length
*	91214	91941: contig of 728 bp in length
*	91942	92021: gap of unknown length
*	92022	92785: contig of 764 bp in length
*	92786	92865: gap of unknown length
*	92866	93548: contig of 683 bp in length

```
            *    93549    93628: gap of unknown length
            *    93629    94374: contig of 746 bp in length
            *    94375    94454: gap of unknown length
            *    94455    95143: contig of 689 bp in length
            *    95144    95223: gap of unknown length
            *    95224    95850: contig of 627 bp in length
            *    95851    95930: gap of unknown length
            *    95931    96666: contig of 736 bp in length
            *    96667    96746: gap of unknown length
            *    96747    97467: contig of 721 bp in length
            *    97468    97547: gap of unknown length
            *    97548    98374: contig of 827 bp in length
            *    98375    98454: gap of unknown length
            *    98455    99170: contig of 716 bp in length
            *    99171    99250: gap of unknown length
            *    99251    99955: contig of 705 bp in length
            *    99956   100035: gap of unknown length
            *   100036   100860: contig of 825 bp in length
            *   100861   100940: gap of unknown length
            *   100941   101652: contig of 712 bp in length
            *   101653   101732: gap of unknown length

  Query Match              3.7%;  Score 153.2;  DB 41;  Length 118474;
  Best Local Similarity   98.1%;  Pred. No. 2.2e-34;
  Matches  155;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gtgtgcctgactgttttgtaggtgtaaggaggggcgtggccaaatagtttttggtatacg 60
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 103315 GTGTGCGTGACTGTTTTGTAGGTGTAAGGAGGGGCGTGGCCAAATAGTTTTTGGTATACG 103374

Qy     61 gatagaatttggatgaaaaataaaacgaaatcaaaacatttttcaaaagcgtggaagttt 120
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 103375 GATAGAAATTGGATGAAAAATAAAACGAAATCAAAACATTTTTCAAAAGCGTGGAAGTTT 103434

Qy    121 tggccggcttgtgggcatggcaaaacgttttttggcta 158
          |||||||||||||||||||||||||||||||| ||
Db 103435 TGGCCGGCTTGTGGGCATGGCAAAACGTTTTTTGGGTA 103472


RESULT  11
AC009255/c
LOCUS       AC009255   109699 bp    DNA             HTG       10-DEC-1999
DEFINITION  Drosophila melanogaster chromosome 2 clone BACR39E09 (D1043)
            RPCI-98 39.E.9 map 42A-42A strain y; cn bw sp, *** SEQUENCING IN
            PROGRESS ***, 103 unordered pieces.
ACCESSION   AC009255
VERSION     AC009255.3  GI:6554247
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 109699)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
            Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
            Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 109699)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Dec 10, 1999 this sequence version replaced gi:5734680.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 103 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1199: contig of 1199 bp in length
            *     1200     1279: gap of unknown length
            *     1280     1997: contig of 718 bp in length
            *     1998     2077: gap of unknown length
            *     2078     2686: contig of 609 bp in length
            *     2687     2766: gap of unknown length
            *     2767     3839: contig of 1073 bp in length
            *     3840     3919: gap of unknown length
            *     3920     4568: contig of 649 bp in length
            *     4569     4648: gap of unknown length
            *     4649     5320: contig of 672 bp in length
            *     5321     5400: gap of unknown length
            *     5401     6335: contig of 935 bp in length
            *     6336     6415: gap of unknown length
            *     6416     7390: contig of 975 bp in length
            *     7391     7470: gap of unknown length
            *     7471     8350: contig of 880 bp in length
            *     8351     8430: gap of unknown length
            *     8431     9510: contig of 1080 bp in length
            *     9511     9590: gap of unknown length
            *     9591    10436: contig of 846 bp in length
            *    10437    10516: gap of unknown length
            *    10517    11243: contig of 727 bp in length
            *    11244    11323: gap of unknown length
            *    11324    11946: contig of 623 bp in length
            *    11947    12026: gap of unknown length
            *    12027    12701: contig of 675 bp in length
            *    12702    12781: gap of unknown length
            *    12782    13062: contig of 281 bp in length
            *    13063    13142: gap of unknown length
            *    13143    14057: contig of 915 bp in length
            *    14058    14137: gap of unknown length
            *    14138    14805: contig of 668 bp in length
            *    14806    14885: gap of unknown length
            *    14886    15334: contig of 449 bp in length
            *    15335    15414: gap of unknown length
            *    15415    16203: contig of 789 bp in length
            *    16204    16283: gap of unknown length
            *    16284    17032: contig of 749 bp in length
            *    17033    17112: gap of unknown length
            *    17113    18667: contig of 1555 bp in length
            *    18668    18747: gap of unknown length
            *    18748    19961: contig of 1214 bp in length
            *    19962    20041: gap of unknown length
            *    20042    20778: contig of 737 bp in length
            *    20779    20858: gap of unknown length
            *    20859    21803: contig of 945 bp in length
            *    21804    21883: gap of unknown length
            *    21884    22739: contig of 856 bp in length
            *    22740    22819: gap of unknown length
            *    22820    23590: contig of 771 bp in length
            *    23591    23670: gap of unknown length
            *    23671    24575: contig of 905 bp in length
            *    24576    24655: gap of unknown length
            *    24656    25422: contig of 767 bp in length
```

```
*    25423    25502: gap of unknown length
*    25503    26176: contig of 674 bp in length
*    26177    26256: gap of unknown length
*    26257    26758: contig of 502 bp in length
*    26759    26838: gap of unknown length
*    26839    27932: contig of 1094 bp in length
*    27933    28012: gap of unknown length
*    28013    29308: contig of 1296 bp in length
*    29309    29388: gap of unknown length
*    29389    30496: contig of 1108 bp in length
*    30497    30576: gap of unknown length
*    30577    31728: contig of 1152 bp in length
*    31729    31808: gap of unknown length
*    31809    33080: contig of 1272 bp in length
*    33081    33160: gap of unknown length
*    33161    34333: contig of 1173 bp in length
*    34334    34413: gap of unknown length
*    34414    35478: contig of 1065 bp in length
*    35479    35558: gap of unknown length
*    35559    36465: contig of 907 bp in length
*    36466    36545: gap of unknown length
*    36546    37699: contig of 1154 bp in length
*    37700    37779: gap of unknown length
*    37780    38966: contig of 1187 bp in length
*    38967    39046: gap of unknown length
*    39047    40514: contig of 1468 bp in length
*    40515    40594: gap of unknown length
*    40595    41900: contig of 1306 bp in length
*    41901    41980: gap of unknown length
*    41981    43438: contig of 1458 bp in length
*    43439    43518: gap of unknown length
*    43519    44606: contig of 1088 bp in length
*    44607    44686: gap of unknown length
*    44687    46717: contig of 2031 bp in length
*    46718    46797: gap of unknown length
*    46798    48201: contig of 1404 bp in length
*    48202    48281: gap of unknown length
*    48282    49781: contig of 1500 bp in length
*    49782    49861: gap of unknown length
*    49862    51105: contig of 1244 bp in length
*    51106    51185: gap of unknown length
*    51186    52319: contig of 1134 bp in length
*    52320    52399: gap of unknown length
*    52400    54918: contig of 2519 bp in length
*    54919    54998: gap of unknown length
*    54999    56678: contig of 1680 bp in length
*    56679    56758: gap of unknown length
*    56759    59360: contig of 2602 bp in length
*    59361    59440: gap of unknown length
*    59441    62856: contig of 3416 bp in length
*    62857    62936: gap of unknown length
*    62937    66768: contig of 3832 bp in length
*    66769    66848: gap of unknown length
*    66849    70379: contig of 3531 bp in length
*    70380    70459: gap of unknown length
*    70460    78725: contig of 8266 bp in length
*    78726    78805: gap of unknown length
*    78806    79327: contig of 522 bp in length
*    79328    79407: gap of unknown length
*    79408    80062: contig of 655 bp in length
*    80063    80142: gap of unknown length
*    80143    80637: contig of 495 bp in length
*    80638    80717: gap of unknown length
*    80718    81225: contig of 508 bp in length
*    81226    81305: gap of unknown length
*    81306    81994: contig of 689 bp in length
*    81995    82074: gap of unknown length
*    82075    82836: contig of 762 bp in length
*    82837    82916: gap of unknown length
*    82917    83557: contig of 641 bp in length
*    83558    83637: gap of unknown length
*    83638    83920: contig of 283 bp in length
*    83921    84000: gap of unknown length
*    84001    84514: contig of 514 bp in length
*    84515    84594: gap of unknown length
*    84595    85404: contig of 810 bp in length
*    85405    85484: gap of unknown length
*    85485    86089: contig of 605 bp in length
*    86090    86169: gap of unknown length
*    86170    86848: contig of 679 bp in length
*    86849    86928: gap of unknown length
*    86929    87211: contig of 283 bp in length
*    87212    87291: gap of unknown length
*    87292    87990: contig of 699 bp in length
*    87991    88070: gap of unknown length
*    88071    88629: contig of 559 bp in length
*    88630    88709: gap of unknown length
*    88710    89373: contig of 664 bp in length
*    89374    89453: gap of unknown length
*    89454    89767: contig of 314 bp in length
*    89768    89847: gap of unknown length
*    89848    90407: contig of 560 bp in length
*    90408    90487: gap of unknown length
*    90488    90961: contig of 474 bp in length
*    90962    91041: gap of unknown length

  Query Match              3.3%;  Score 137;  DB 42;  Length 109699;
  Best Local Similarity   58.7%;  Pred. No. 1.9e-29;
  Matches  215;  Conservative    0;  Mismatches   86;  Indels   65;  Gaps    1;

Qy     415 cgaaccagtgttctcccgtggacggagggaatgcagcttatggatgcgtttcgtgcgccg 474
           |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   83918 CGAACCAGTGTTCTCCCGTGGACGGAGGGAATTCAGCTTATGGATGCGTTTCGTGCGCCG 83859

Qy     475 ctacacgaagtttttaaattgcttgaaattgtgcgcaatcaccagagcagcgaaaa---- 530
           |||||||||||||||||||||||||||||||||||||||||||||| | |||||||||
Db   83858 CTACACGAAGTTTTTAAATTGCTTGAAATTGTGCGCAATCACCANACCAGCGAAAAGTGA 83799

Qy     531 ------------------------------------------------------------ 530

Db   83798 GTACATCCAAGGGGTCTTCCAGTCTCTCAACAACACTGATTTACTTAATACTATGCGACA 83739

Qy     531 -caaacgtaccctggagcacaactgcctacatgtagacaacgtaaagcgcggaacacacg 589
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   83738 GCAAACGTACCCTGGAGCACAACTGCCTACATGTAGACAACGTAAAGCGCGGAACACACG 83679

Qy     590 ggcagctggaccagatctttccggagtatggctgcctgctgctctcgcccgccaacctgt 649
           ||||||||||||||||||||||||||||||||||||||||
Db   83678 GGCAGCTGGACCAGATCTTTCCGGAGTATGGCTGCCTGCTGNNNNNNNNNNNNNNNNNNN 83619

Qy     650 ggacgcagaactctcagaactttactcgggacacaaacatcctgaacacgatatttcagt 709

Db   83618 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 83559

Qy     710 accata 715
             | |
Db   83558 NTAAAA 83553


RESULT  12
AC017232/c
LOCUS       AC017232    10844 bp    DNA             HTG       09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC017232
VERSION     AC017232.1  GI:6553754
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 10844)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
```

```
          Rockville, MD, USA
COMMENT   This sequence was identified as CDM:10209962 by the submitter.
          For more information on this record e-mail to fly@celera.com.
          * NOTE: This is a 'working draft' sequence.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .10844
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     3202 a    2155 c    2105 g    3382 t
ORIGIN


  Query Match             1.5%;  Score 62.4;  DB 42;  Length 10844;
  Best Local Similarity   75.3%;  Pred. No. 6e-07;
  Matches  110;  Conservative    0;  Mismatches   21;  Indels   15;  Gaps    2;

Qy       8 tgactgttttgtaggtgtaaggaggggcgtggccaaatagtttttggtatacggatagaa 67
           ||  |||||||||||||||||||||||||||| |||  |||||||||||| || ||||
Db    5920 TGGGTGTTTTGTAGGTGTAAGGAGGGGCGTGGCAAAAGTGTTTTTGGTATACCGACAGAA 5861

Qy      68 tttggatga---------aaaataaaacgaaatcaaaacatttttcaaaa------gcgt 112
            ||||   |         ||||  ||   ||||||||||||||||||||      ||||
Db    5860 ATTGGTAAAACAAACAATAAAACCAAGAAAAATCAAAACATTTTTCAAAAGTGTGGGCGT 5801

Qy     113 ggaagttttggccggcttgtgggcat 138
           | ||||||||  |||||||||| |
Db    5800 AGCAGTTTTGGGAGGCTTGTGGGCGT 5775


RESULT  13
AE003066/c
LOCUS       AE003066    28380 bp    DNA             INV       22-MAR-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013385325, complete
            sequence.
ACCESSION   AE003066
VERSION     AE003066.1  GI:7289027
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 28380)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 28380)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
FEATURES             Location/Qualifiers
     source          1. .28380
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
     mRNA            complement(join(5607. .6323,6668. .6957,7095. .7257))
                     /gene="CG13865"
                     /product="CT33380"
                     /db_xref="FLYBASE:FBan0013865"
                     /db_xref="FLYBASE:FBgn0039965"
     gene            complement(<5607. .>7257)
                     /gene="CG13865"
                     /db_xref="FLYBASE:FBan0013865"
                     /db_xref="FLYBASE:FBgn0039965"
     CDS             complement(join(5913. .6323,6668. .6957,7095. .7194))
                     /gene="CG13865"
                     /note="CG13865 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0013865"
                     /db_xref="FLYBASE:FBgn0039965"
                     /protein_id="AAF45339.1"
                     /db_xref="GI:7289028"
                     /translation="MAYDLKKESDVKEYVEKLGVEYRFGCYSEKKPEACHLLGDYLEG
                     IKKDFEKASKVYKSTCDDYGYAKSCYKYGNYSFLGKGKSGSKGNPQVAYEYYEKGCNL
                     NDSDACLHSGLLLVSKSMPREIDWNVPKGLEFLTKSCDLNNATACFYLSGMHISGVQK
                     KADQSAVTASSGSGTSSPPAGQPPLKDSDYIVLKDMKKAFQFAHKACELRNMYACANL
                     SQMYARGDGIEKNEKEAEKYKKLALEMQDEVKKQHDTLGFQQGVGMPN"
     mRNA            join(7547. .7733,7795. .>8399)
                     /gene="CG17515"
                     /product="CT33382"
                     /db_xref="FLYBASE:FBan0017515"
                     /db_xref="FLYBASE:FBgn0039966"
     gene            <7547. .>8399
                     /gene="CG17515"
                     /db_xref="FLYBASE:FBan0017515"
                     /db_xref="FLYBASE:FBgn0039966"
     CDS             join(7670. .7733,7795. .8399)
                     /gene="CG17515"
                     /note="CG17515 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0017515"
                     /db_xref="FLYBASE:FBgn0039966"
                     /protein_id="AAF45341.1"
                     /db_xref="GI:7289030"
                     /translation="MSSSRTRNGPTLNFKAVLLGEGCVGKTSLVLRYMEDRFNAQHLS
                     TLQASFVSRKMSLEDGRRAQLNIWDTAGQERFHALGPIYYRGSDGALLVYDITDRDSF
                     QKVKSWVRELRQMRGTEIALIIVGNKTDLEEQRAVTHDEALQYARTVGAQYVETSAKE
                     NEGVAELFELLTQLMLEQLSQRQPDASPLRLQNPDTDNLNNSDDSEAPDPGDPAGQRS
                     CCGI"
     mRNA            join(<24856. .25554,25613. .26131,26612. .27888)
                     /gene="stnA"
                     /note="Nucleotide sequence of the Celera sequence differs
                     from the published sequence for this transcript."
                     /product="CT33379"
                     /db_xref="FLYBASE:FBan0012500"
```

```
                     /db_xref="FLYBASE:FBgn0016976"
                     /evidence=not_experimental
     gene            <24856. .>27888
                     /gene="stnA"
                     /note="CG12500"
                     /map="20A4-20A5"
                     /db_xref="FLYBASE:FBan0012500"
                     /db_xref="FLYBASE:FBgn0016976"
                     /evidence=not_experimental
     CDS             join(24856. .25554,25613. .25663)
                     /gene="stnA"
                     /note="stnA gene product; Nucleotide sequence of the
                     Celera sequence differs from the published sequence for
                     this transcript"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0012500"
                     /db_xref="FLYBASE:FBgn0016976"
                     /evidence=not_experimental
                     /protein_id="AAF45340.1"
                     /db_xref="GI:7289029"
                     /translation="MLKLPKGLKKKKKKSKKDQELFTEEELEQYKRDLKAKQEAAATK
                     SDAGESDGASSDVEAHHEPVAFNSGLGSGSSSSILNAQQQLSDQNQGAAGGDEEWAKF
                     KALTSGVDSILHKTQDELDRIKKESFYQRLPSAAEKKKQKEEEAARLEAEQQEREKQR
                     LGQIEAQRDKLAEAVVQLSESEEEAGDYEADDIFATDYIEAITSGELQLAVVPDSPVL
                     AEDGPDPFDTAYAEKPLRFYPVAWTESTLSP"
BASE COUNT     6404 a   4440 c   4208 g   6616 t   6712 others
ORIGIN


  Query Match             1.5%;  Score 62.4;  DB 34;  Length 28380;
  Best Local Similarity   75.3%;  Pred. No. 7.5e-07;
  Matches  110;  Conservative    0;  Mismatches   21;  Indels   15;  Gaps    2;

Qy       8 tgactgttttgtaggtgtaaggaggggcgtggccaaatagtttttggtatacggatagaa 67
           ||  ||||||||||||||||||||||||||||| |||  ||||||||||||| || ||||
Db    9428 TGGGTGTTTTGTAGGTGTAAGGAGGGGCGTGGCAAAAGTGTTTTTGGTATACCGACAGAA 9369

Qy      68 tttggatga---------aaaataaaacgaaatcaaaacatttttcaaaa------gcgt 112
           ||||   |          ||||  ||   ||||||||||||||||||||      ||||
Db    9368 ATTGGTAAAACAAACAATAAAACCAAGAAAAATCAAAACATTTTTCAAAAGTGTGGGCGT 9309

Qy     113 ggaagttttggccggcttgtgggcat 138
           | ||||||||  ||||||||||| |
Db    9308 AGCAGTTTTGGGAGGCTTGTGGGCGT 9283


RESULT  14
AC017996
LOCUS       AC017996    16047 bp    DNA             HTG       09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC017996
VERSION     AC017996.1  GI:6553194
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 16047)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212940 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .16047
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     4912 a   3114 c   3241 g   4780 t
ORIGIN


  Query Match             1.5%;  Score 61.4;  DB 42;  Length 16047;
  Best Local Similarity   75.5%;  Pred. No. 1.3e-06;
  Matches  105;  Conservative    0;  Mismatches   26;  Indels    8;  Gaps    2;

Qy       8 tgactgttttgtaggtgtaaggaggggcgtggccaaatagtttttggtatacggatagaa 67
           ||   ||||||||||||||||||||||||||||||||||| |||||| |||||| ||| |||
Db    6827 TGGGCGTTTTGTAGGTGTAAGGAGGGGCGTGGCCAAATTGTTTTTAGTATACCGATTGAA 6886

Qy      68 tttggatga--aaaataaaacgaaatcaaaacatttttcaaa------agcgtggaagtt 119
           ||  |  |  | ||   |  ||||||||||| ||||||||      |||||| ||||
Db    6887 ATTCAAACAATATAACTGAGAAAAATCAAAACACTTTTCAAAAGTGTGAGCGTGGCAGTT 6946

Qy     120 ttggccggcttgtgggcat 138
           |||| | | ||||| | |
Db    6947 TTGGTCTGTTTGTGGACGT 6965


RESULT  15
AE003781
LOCUS       AE003781   311786 bp    DNA             INV       22-MAR-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386030 section 1
            of 3, complete sequence.
ACCESSION   AE003781 AE002725
VERSION     AE003781.1  GI:7302115
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 311786)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
```

```
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 311786)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
FEATURES             Location/Qualifiers
     source          1. .311786
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
     mRNA            join(16241. .16424,17650. .17976,18828. .19049,19106. .>19277)
                     /gene="CG3305"
                     /product="CT11115"
                     /db_xref="FLYBASE:FBan0003305"
                     /db_xref="FLYBASE:FBgn0032949"
     gene            <16241. .>19277
                     /gene="CG3305"
                     /db_xref="FLYBASE:FBan0003305"
                     /db_xref="FLYBASE:FBgn0032949"
     CDS             join(16358. .16424,17650. .17976,18828. .19049,19106. .>19277)
                     /gene="CG3305"
                     /note="CG3305 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0003305"
                     /db_xref="FLYBASE:FBgn0032949"
                     /protein_id="AAF57215.1"
                     /db_xref="GI:7302116"
                     /translation="MFANKLLTCSALLIMLFLLSSTVFSQKLIGLPRVGKNVSDLFLE
                     QANPLTSSSSTISTSSTTTEKPITTRSTSTITPRSTTPSTESPSTTAPVISTTIAPQP
                     YPQPSIGAWNTSCIMLQMAAQLNFTYEAREGNFTTGLYNIPSNASVEDAECKSQTTQF
                     IHLIWGPETSKQSLIMYFNKSNDTTVLSFMQIHLALLPEDFPDAKENQTVQLITRSDG
                     AFKTPENMSYHCTRVQKINMTETLDAEQLIGWISVSHVQVEAFR"
     mRNA            complement(<20068. .>20397)
                     /gene="CG14465"
                     /product="CT34156"
                     /db_xref="FLYBASE:FBan0014465"
                     /db_xref="FLYBASE:FBgn0032950"
                     /evidence=not_experimental
     mRNA            complement(join(<20068. .20412,21319. .21392,21594. .22485,
                     24480. .24679,29305. .30046,30136. .30164,32793. .>33192))
                     /gene="CG12548"
                     /product="CT34155"
                     /db_xref="FLYBASE:FBan0012548"
                     /db_xref="FLYBASE:FBgn0032952"
     gene            complement(<20068. .>20397)
                     /gene="CG14465"
                     /db_xref="FLYBASE:FBan0014465"
                     /db_xref="FLYBASE:FBgn0032950"
                     /evidence=not_experimental
     gene            complement(<20068. .>33192)
                     /gene="CG12548"
                     /db_xref="FLYBASE:FBan0012548"
                     /db_xref="FLYBASE:FBgn0032952"
     CDS             complement(join(20068. .20412,21319. .21392,21594. .22485,
                     24480. .24679,29305. .30046,30136. .30164,32793. .33192))
                     /gene="CG12548"
                     /note="CG12548 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0012548"
                     /db_xref="FLYBASE:FBgn0032952"
                     /protein_id="AAF57216.1"
                     /db_xref="GI:7302117"
                     /translation="MHIEALNTYSIMAKNKMFPHVNQLKLNMGNIYYSMGIYQKAVKM
                     YRMALDSVPKSLSQLRLKIRENIGILFIRMGSYSDAASSFEFIMTERANIRSSIHLLL
                     CYFALGDVEKVKLAFRRLCDVQTEAIESDMEKHYTNFETSTVFFNLAEQYERSEMHIE
                     ALNTYSIMAKNKMFPHVNQLKLNMGNIYYSMGIYQKAVKMYRMALDSVPKSLSQLRLK
                     IRENIGILFIRMGSYSDAASSFEFIMTERANIRSSIHLLLCYFALGDVEKVKLAFRRL
                     CDVQTEAIESDMESETNIIKLQQQAEPIQQIGETDGYQSLNNEPVTGSVIKAEGGGKL
                     NETVKFAATVKHRYVVQALKKDELAVYRNERRNAVKRSITMIVDLISPFIEDNYNDGY
                     NWCIEIIKTSNLAWLANELELNKALVYLRQNDVHQAIETLQMYDRKSEGSMTASALTN
                     LSFIYISMASHCVNQLHEIGSLKNNAPGLINAGIVELGSHNLILASERFEGALQLQPM
                     NFEARYNLGLVALAQNDYELAEERFELLKEQLMLPSSVQHSHVFYQLAKLQERRLESG
                     LISNFTPGAALQAYLQVVGISASDIDSRLFEKVGSLYEQIQDHQEANQYYNEAYRINM
                     SDIGIASSIGSYYIKLQATERALFYYERAVLADPNDPNLMLRIASCFRNSYLPTKQYL
                     GLFQKIHARFPDNLTCIRALMQVTKSLELGDLYERYASEYTRLQSQQQERQSEQRYQQ
                     QRFPSAAKFSNRLRFLQFDPLGPPAERPKTGRPLQSRDDLGNLEMASHCVNQLHEIGS
                     LKNNAPGLINAGIVELDSHNLILARERFEGALQLQPMNFEARYNLGLVALAQNDYELA
                     EERFELLKEQLMMPSSGHVFCQLQERRISSKQEQRNN"
     CDS             complement(20068. .20397)
                     /gene="CG14465"
                     /note="CG14465 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0014465"
                     /db_xref="FLYBASE:FBgn0032950"
                     /evidence=not_experimental
                     /protein_id="AAF57217.1"
                     /db_xref="GI:7302118"
                     /translation="MASHCVNQLHEIGSLKNNAPGLINAGIVELDSHNLILARERFEG
                     ALQLQPMNFEARYNLGLVALAQNDYELAEERFELLKEQLMMPSSGHVFCQLQERRISS
                     KQEQRNN"
     mRNA            complement(<29291. .>30007)
                     /gene="CG4137"
                     /product="CT5172"
                     /db_xref="FLYBASE:FBan0004137"
                     /db_xref="FLYBASE:FBgn0032951"
                     /evidence=not_experimental
     gene            complement(<29291. .>30007)
                     /gene="CG4137"
                     /db_xref="FLYBASE:FBan0004137"
                     /db_xref="FLYBASE:FBgn0032951"
                     /evidence=not_experimental
     CDS             complement(29291. .30007)
                     /gene="CG4137"
                     /note="CG4137 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0004137"
                     /db_xref="FLYBASE:FBgn0032951"
                     /evidence=not_experimental
                     /protein_id="AAF57218.1"
                     /db_xref="GI:7302119"
                     /translation="MHIEALNTYSIMAKNKMFPHVNQLKLNMGNIYYSMGIYQKAVKM
                     YRMALDSVPKSLSQLRLKIRENIGILFIRMGSYSDAASSFEFIMTERANIRSSIHLLL
                     CYFALGDVEKVKLAFRRLCDVQTEAIESDMESETNIIKLQQQAEPIQQIGETDGYQSL
                     NNEPVTGSVIKAEGGGKLNETVKFAATVKHRYVVQALKKDELAVYRNERRNAVKRSIT
                     MIVDLISPFIEDNYNDGEYD"
     mRNA            complement(<32689. .>33192)
                     /gene="CG15190"
                     /product="CT35106"
                     /db_xref="FLYBASE:FBan0015190"
                     /db_xref="FLYBASE:FBgn0032953"
                     /evidence=not_experimental
     gene            complement(<32689. .>33192)
                     /gene="CG15190"
                     /db_xref="FLYBASE:FBan0015190"
                     /db_xref="FLYBASE:FBgn0032953"
                     /evidence=not_experimental
     CDS             complement(32689. .33192)
                     /gene="CG15190"
                     /note="CG15190 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0015190"
                     /db_xref="FLYBASE:FBgn0032953"
                     /evidence=not_experimental
                     /protein_id="AAF57219.1"
                     /db_xref="GI:7302120"
                     /translation="MHIEALNTYSIMAKNKMFPHVNQLKLNMGNIYYSMGIYQKAVKM
                     YRMALDSVPKSLSQLRLKIRENIGILFIRMGSYSDAASSFEFIMTERANIRSSIHLLL
                     CYFALGDVEKVKLAFRRLCDVQTEAIESDMESETNIIKLQQQAEPIQQCGEIDVHQTT
                     NKTRARL"
     mRNA            join(<36269. .36944,37634. .37921)
                     /gene="CG3549"
                     /note="Transcript structure is suspect."
```

```
                     /product="CT11930"
                     /db_xref="FLYBASE:FBan0003549"
                     /db_xref="FLYBASE:FBgn0032954"
     gene            <36269. .>37921
                     /gene="CG3549"

  Query Match             1.5%;  Score 61.4;  DB 58;  Length 311786;
  Best Local Similarity   75.5%;  Pred. No. 2.7e-06;
  Matches  105;  Conservative    0;  Mismatches   26;  Indels    8;  Gaps    2;

Qy       8 tgactgttttgtaggtgtaaggaggggcgtggccaaatagtttttggtatacggatagaa 67
           ||   ||||||||||||||||||||||||||||||||||| |||||| |||||| ||| |||
Db  163121 TGGGCGTTTTGTAGGTGTAAGGAGGGGCGTGGCCAAATTGTTTTTAGTATACCGATTGAA 163180

Qy      68 tttggatga--aaaataaaacgaaatcaaaacatttttcaaa------agcgtggaagtt 119
            ||  |  |  | ||   |  ||||||||||| ||||||||      ||||||| ||||
Db  163181 ATTCAAACAATATAACTGAGAAAAATCAAAACACTTTTCAAAAGTGTGAGCGTGGCAGTT 163240

Qy     120 ttggccggcttgtgggcat 138
           |||| | | |||||| | |
Db  163241 TTGGTCTGTTTGTGGACGT 163259


Search completed: September  4, 2000, 00:48:39
Job time: 22339 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2000, 00:24:48 ; Search time 4457.97 Seconds
(without alignments)
4083.422 Million cell updates/sec

Title: US-09-332-522B-5
Perfect score: 4128
Sequence: 1 gtgtgcctgactgttttgta.....tgactaaatgttaaaataaa 4128

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: em_est35:*
83: em_est36:*
84: em_est37:*
85: gb_est48:*
86: gb_est49:*
87: gb_est50:*
88: gb_est51:*
89: gb_est52:*
90: gb_est53:*
91: gb_est54:*
92: gb_est55:*
93: gb_gss1:*
94: gb_gss2:*
95: gb_gss3:*
96: gb_gss4:*
97: em_gss1:*
98: em_gss2:*
99: em_gss3:*
100: em_gss4:*
101: gb_gss5:*
102: gb_gss6:*
103: gb_gss7:*
104: gb_gss8:*
105: gb_gss9:*
106: em_gss5:*
107: em_gss6:*
108: em_gss7:*
109: em_gss8:*
110: em_gss9:*
111: em_gss10:*
112: em_gss11:*
113: gb_gss10:*
114: gb_gss11:*
115: em_gss12:*
116: gb_gss12:*

```
117:  gb_gss13:*
118:  gb_gss14:*
119:  gb_gss15:*
120:  gb_gss16:*
121:  gb_gss17:*
122:  gb_gss18:*
123:  gb_gss19:*
124:  em_gss13:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	363.4	8.8	754	122	CNS0116H	AL099875 Drosophil
	2	321.6	7.8	328	46	AI945842	AI945842 bs17e03.y
	3	99.6	2.4	390	69	AW240273	AW240273 up29g10.y
	4	80.4	1.9	546	79	AW654475	AW654475 103951 MA
	5	73.2	1.8	399	88	R55122	R55122 yg87f07.r1
	6	61.4	1.5	561	27	AA498678	AA498678 vh33d06.r
	7	61.2	1.5	432	21	AA106247	AA106247 mm18g04.r
	8	60.4	1.5	647	23	AA254928	AA254928 mz84c05.r
	9	60.2	1.5	611	47	AL043581	AL043581 DKFZp434F
	10	51.8	1.3	333	25	AA354156	AA354156 EST62373
	11	50	1.2	1101	122	CNS001UZ	AL061599 Drosophil
c	12	49.4	1.2	1201	123	CNS0163J	AL106249 Drosophil
c	13	48.4	1.2	1101	122	CNS00AX2	AL056044 Drosophil
	14	48	1.2	1101	122	CNS00ERH	AL069558 Drosophil
	15	47.2	1.1	982	122	CNS00EQ8	AL069513 Drosophil
	16	47	1.1	529	79	AW632482	AW632482 92442 MAR
c	17	46.4	1.1	424	79	AW662299	AW662299 hi23c06.x
c	18	46.4	1.1	792	47	AL040564	AL040564 DKFZp434J
	19	46.4	1.1	1101	122	CNS00BMG	AL057362 Drosophil
	20	46.2	1.1	360	81	D67504	D67504 CELK053DYF
	21	46.2	1.1	440	22	AA195538	AA195538 zr32a07.r
	22	45.4	1.1	517	79	AW656527	AW656527 108828 MA
	23	45.2	1.1	940	122	CNS00CF0	AL059197 Drosophil
c	24	45	1.1	959	122	CNS00Z2S	AL097150 Drosophil
	25	44.6	1.1	503	81	C89382	C89382 C89382 Mous
	26	44.4	1.1	312	72	AW417103	AW417103 53063 MAR
	27	44.4	1.1	1028	122	CNS00LD8	AL067969 Drosophil
	28	44.2	1.1	986	122	CNS00KT1	AL077996 Drosophil
	29	43.8	1.1	255	80	C18401	C18401 C18401 Huma
	30	43.8	1.1	326	25	AA378263	AA378263 EST90950
	31	43.8	1.1	448	47	AL045225	AL045225 DKFZp434H
	32	43.8	1.1	511	22	AA195161	AA195161 zr35g06.r
c	33	43.8	1.1	634	47	AL043582	AL043582 DKFZp434F
	34	43.8	1.1	832	122	CNS007XU	AL050972 Drosophil
c	35	43.4	1.1	561	36	AI218269	AI218269 qh28f02.x
c	36	43.4	1.1	623	72	AW439423	AW439423 hb85c05.x
c	37	43.2	1.0	429	64	AW125000	AW125000 UI-M-BH2.
	38	43.2	1.0	446	22	AA217968	AA217968 mv57c03.r
	39	43.2	1.0	449	20	AA049660	AA049660 mj37a06.r
	40	43.2	1.0	491	29	AA625370	AA625370 af71f01.r
	41	43.2	1.0	491	36	AI154968	AI154968 ud82b09.r
	42	43.2	1.0	492	26	AA422332	AA422332 vf45b11.r
	43	43.2	1.0	515	32	AA870601	AA870601 vq23g05.r
	44	43.2	1.0	1101	122	CNS00LHX	AL068239 Drosophil
c	45	43	1.0	465	35	AI145323	AI145323 UI-R-BT0-

ALIGNMENTS

```
RESULT   1
CNS0116H/c
LOCUS       CNS0116H      754 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN05M24 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL099875
VERSION     AL099875.1  GI:5611486
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 754)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .754
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN05M24"
                     /note="end : T7"
BASE COUNT      198 a    129 c    183 g    217 t     27 others
ORIGIN


  Query Match             8.8%;  Score 363.4;  DB 122;  Length 754;
  Best Local Similarity   94.2%;  Pred. No. 4.8e-98;
  Matches  359;  Conservative   15;  Mismatches    7;  Indels    0;  Gaps    0;

Qy     150 ttttggctatccgttaatcaacataccgttgcccgggacaatacccaccaagatcgttgt 209
           |||  |||||||:|||||||| ||||||||||||||:| ||||||||||||||||||||
Db     438 TTTCAGCTATCCRTTAATCAAAATACCGTTGCCCGGRAAAATACCCACCAAGATCGTTGT 379

Qy     210 accctacgaaactggatccggatcgctgtcatggcactctcttaatacatcctcgactac 269
           |||||||:||||||||||||||:|||||||||||||:||||||||||||||||||||||
Db     378 ACCCTACRAAACTGGATCCGGATCGYTGTCATGGCACTCTMTTAATACATCCTCGACTAC 319

Qy     270 accgcaggaaccgcacccttccggcgaaccctggccccccgaaccacaggtactcaatag 329
           |||||||||||||:|||||||||||||||||||||||::|||||||:||||:|||||
Db     318 ACCGCAGGAACCGCMCCCTTCCGGCGAACCCTGGCCCCCCVMACCACAGKTACTMAATAG 259

Qy     330 cagtaccacggaccgcagcccgcctccccttctgccctgggcgcagagcagccccgccctt 389
           |||||||||||||||||||||||||||||||||||||||||||||: : :||| ||||||:||
Db     258 CAGTACCACGGACCGCAGCCCGCCTCCCCTTCTGCCCTGGGCKTWTWGCACCCCCGCYTT 199

Qy     390 tttctacgtccagcagattactctgcgaaccagtgttctcccgtggacggagggaatgca 449
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     198 TTTCTACGTCCAGCAGATTACTCTGCGAACCAGTGTTCTCCCGTGGACGGAGGGAATGCA 139

Qy     450 gcttatggatgcgtttcgtgcgccgctacacgaagtttttaaattgcttgaaattgtgcg 509
           ||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db     138 GCTTATGGATGCGTTTCGTGCGCCGCTACACGAAKTTTTTAAATTGCTTGAAATTGTGCG 79

Qy     510 caatcaccagagcagcgaaaa 530
           |||||||||||||||||||||
Db      78 CAATCACCAGAGCAGCGAAAA 58


RESULT   2
AI945842
LOCUS       AI945842      328 bp    mRNA            EST       17-AUG-1999
DEFINITION  bs17e03.y1 Drosophila melanogaster adult testis library Drosophila
            melanogaster cDNA clone bs17e03 5', mRNA sequence.
```

```
ACCESSION   AI945842
VERSION     AI945842.1  GI:5736240
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 328)
  AUTHORS   Andrews, J., Bouffard, G. and Oliver, B.
  TITLE     Drosophila melanogaster testis expressed sequence tags
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Brian Oliver
            Laboratory of Cellular and Developmental Biology
            NIDDK, National Institutes of Health
            6 Center Drive MSC 2715, Bldg 6, Rm B1-13, Bethesda, MD 20892 USA
            Fax: (301) 496 5239
            Email: oliver@helix.nih.gov,
            http://www.niddk.nih.gov/intram/people/boliver.htm
            Tissue isolation and library construction performed at the National
            Institute of Diabetes and Digestive and Kidney Diseases, NIH (see
            http://www.niddk.nih.gov/intram/people/boliver.htm). DNA sequencing
            and analyses performed by National Institutes of Health Intramural
            Sequencing Center (NISC; see http://www.nisc.nih.gov). Chromatogram
            data were analyzed and evaluated for high quality using the ted
            program (Gleeson T and Hillier L, 1991).
            Plate: 17  row: e  column: 03
            Seq primer: M13RP1 reverse primer (ABI).
FEATURES             Location/Qualifiers
     source          1. .328
                     /organism="Drosophila melanogaster"
                     /strain="y[*] w[67c1]/Y"
                     /db_xref="taxon:7227"
                     /clone="bs17e03"
                     /clone_lib="Drosophila melanogaster adult testis library"
                     /sex="male"
                     /dev_stage="1-5 day adult"
                     /lab_host="SOLR (Stratagene)"
                     /note="Organ: testis; Vector: pBlueScript SK (Stratagene);
                     Site_1: EcoR I; Site_2: Xho I; Testes dissected from 1-5
                     day adult y[*] w[67c1]/Y males raised at 25oC.  RNA
                     isolated using Trizol (Life Technologies) and a single
                     round of Poly(A)+ selection using Oligotex (Qiagen).  cDNA
                     library constructed using Stratagene ZAP-cDNA syntesis
                     kit.  Oligo dT-primed, size fractionated ~1-6 kb, and
                     directionally cloned at EcoRI and XhoI in Uni-ZAP XR.
                     Following a single round of amplification pBlueScript SK
                     phagemids were mass excised.  A distribution channel for
                     clones is being sought, but not currently available.
                     Requests for clones cannot be honored."
BASE COUNT       95 a     77 c     91 g     65 t
ORIGIN
```

```
  Query Match             7.8%;  Score 321.6;  DB 46;  Length 328;
  Best Local Similarity   98.8%;  Pred. No. 1.4e-85;
  Matches  324;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy   2765 gcgccaacggccgcgtagaattgtgggaatcccctgcgggattgcttaagtgtgcatacc 2824
          || | | ||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      1 GCACGAGCGGCCGCGTAGAATTGTGGGAATCCCCTGCCGGATTGCTTAAGTGTGCATACC 60

Qy   2825 aggaagacgcgaagagaaaccagggtataacccacatccacctgaacggcgatcgagtga 2884
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGGAAGACGCGAAGAGAAACCAGGGTATAACCCACATCCACCTGAACGGCGATCGAGTGA 120

Qy   2885 ttgtggcgcgtcttaatggccgactagatttttaccgcttagagacgtactacaagggga 2944
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTGTGGCGCGTCTTAATGGCCGACTAGATTTTTACCGCTTAGAGACGTACTACAAGGGGA 180

Qy   2945 agcaaatcgactggggttttacctcggcttacaggagaactcatgttcgaactggatcca 3004
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AGCAAATCGACTGGGGTTTTACCTCGGCTTACAGGAGAACTCATGTTCGAACTGGATCCA 240
```

```
Qy   3005 ctggaagcctgggattaatgttgcagcagcagcgctgtcagcaagaagcatcccagaaga 3064
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTGGAAGCCTGGGATTAATGTTGCAGCAGCAGCGCTGTCAGCAAGAAGCATCCCAGAAGA 300

Qy   3065 ccaccaaggaggaaatgaaaatcacatt 3092
          |||||||||||||||||||||||||||
Db    301 CCACCAAGGAGGAAATGAAAATCACATT 328
```

```
RESULT   3
AW240273
LOCUS       AW240273      390 bp    mRNA            EST       14-DEC-1999
DEFINITION  up29g10.y1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:2655810 5'
            similar to TR:P97260 P97260 SREBP CLEAVAGE ACTIVATING PROTEIN. ;,
            mRNA sequence.
ACCESSION   AW240273
VERSION     AW240273.1  GI:6574096
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 390)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 8, 1999 this sequence version replaced gi:4388424.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            MGI:1033958
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .390
                     /organism="Mus musculus"
                     /strain="FVB-3"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2655810"
                     /clone_lib="NCI_CGAP_Mam2"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT       79 a    116 c    105 g     88 t      2 others
ORIGIN
```

```
  Query Match             2.4%;  Score 99.6;  DB 69;  Length 390;
  Best Local Similarity   58.4%;  Pred. No. 1.2e-18;
  Matches  174;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy   1208 tggaaaatagcttggtgatcacaaagagcgtagtctcaatggacgagacattcgacgtga 1267
          | || ||    |||||| |||| |||   || || ||||     | |    |||| ||||
Db      2 TAGAGAACGTGTTGGTGCTCACCAAGTCAGTGGTATCAACTCCAGTGGACCTCGAGGTGA 61

Qy   1268 agatccgcgtggcgcaggctcttagcaaggagggttggcatatatccaagactcttttga 1327
          || | ||  | || || |   | || |  ||| | |||   ||    ||||       ||
Db     62 AGCTTCGGATTGCACAAGGCTTGAGTAGTGAGAGCTGGTCCATCATGAAGAACGCGGCGA 121

Qy   1328 cggagataacaattttgacaattggtcttgctactttcgtgcccgtcatccaggagtttt 1387
```

```
Db    122 CCGAGCTGGGCATCATCCTCATTGGCTACTTCACCCTCGTGCCTGCTATCCAGGAGTTCT 181

Qy   1388 gtatctttgccatagtcggcttgctttccgattttatgctacagatgctgctcttctcaa 1447
Db    182 GCCTCTTTGCTGTTGTGGGCCTGGTGTCTGACTTCTTCCTCCAGATGCTGTTCTTCACCA 241

Qy   1448 caatactggccatgaacattaagcggaccgagtatacggcggaggccaagcaccttcc 1505
Db    242 CTGTCCTGTCGATCGACATTCGCCGGATGGAGCTAGCAGACCTAAACAAGCGGCTGCC 299


RESULT   4
AW654475
LOCUS       AW654475      546 bp    mRNA            EST       05-APR-2000
DEFINITION  103951 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW654475
VERSION     AW654475.1  GI:7420301
KEYWORDS    EST.
SOURCE      Bos taurus.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 546)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and
            Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     On May 6, 1998 this sequence version replaced gi:3114943.
            Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 88  row: F  column: 6
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .546
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 1BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from lymph node, ovary,
                     fat, hypothalamus, and pituitary."
BASE COUNT       99 a    187 c    165 g     95 t
ORIGIN


  Query Match             1.9%;  Score 80.4;  DB 79;  Length 546;
  Best Local Similarity   53.3%;  Pred. No. 8.1e-13;
  Matches  196;  Conservative    0;  Mismatches  166;  Indels    6;  Gaps    1;

Qy   3109 gcccatcagcagccaatcacatgcatgcaggtcgttaacgacatggtttttcactggcagc 3168
Db     32 GCACACCAGAAGCCCATCACAGCCCTGAAGGCCGCTGCCGGGCGCCTCGTGACTGGCAGC 91

Qy   3169 caggatcacaccctcaaggtgtattgcctcaataagtcggatgttgagtatacgctccac 3228
Db     92 CAGGACCACACACTGAGAGTGTTCCGCCTGGAGGATTCGTGCTGCCTCTTCACCCTGCAG 151

Qy   3229 ggtcactgtgggcctgtaacctgtctctttgtggatcgctggcaacctggcacaggggg 3288
```

```
Db    152 GGCCACTCGGGGGCCATCACGACCGTGTATATTGA------CCAGACCATGGTGCTGGCC 205

Qy   3289 tctgggtcccaggacggcctgctctgcgtatgggatctgttcacgggagcctgcatgtat 3348
Db    206 AGCGGAGGCCAAGACGGGGCCATCTGCCTGTGGGACGTGCTGACCGGCAGCCGGGTCAGC 265

Qy   3349 aatatacaagctcacgacggagccgtcagctgcctggcctgtgcgcccagttacgtaatc 3408
Db    266 CACATGTTCGCTCACCGTGGGGATGTCACCTCCCTCACCTGCACCACCTCGTGCGTCATC 325

Qy   3409 tcgctaggcacggacgagaggatttgcgtatgggaacgatttcagggaaacctgttgact 3468
Db    326 AGCAGTGGCCTGGACGACCTCATCAGCATCTGGGACCGCAGCACGGGCATCAAGCTCTAC 385

Qy   3469 accatcaa 3476
Db    386 TCCATCCA 393


RESULT   5
R55122
LOCUS       R55122        399 bp    mRNA            EST       22-MAY-1995
DEFINITION  yg87f07.r1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:40365 5' similar to SP:D2013.8 CE00933 ;, mRNA sequence.
ACCESSION   R55122
VERSION     R55122.1  GI:824458
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 399)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On May 9, 1995 this sequence version replaced gi:803832.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 3115
            High quality sequence stops: 302 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 3115   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 302.
FEATURES             Location/Qualifiers
     source          1. .399
                     /organism="Homo sapiens"
                     /db_xref="GDB:412906"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:40365"
                     /clone_lib="Soares infant brain 1NIB"
                     /sex="female"
                     /dev_stage="73 days post natal"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                     I; Site_2: Hind III; 1st strand cDNA was primed with a Not
                     I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Hind III adaptors
                     (Pharmacia), digested with Not I and directionally cloned
                     into the Not I and Hind III sites of the Lafmid BA vector.
                     Library went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
```

```
BASE COUNT       72 a    120 c    105 g     90 t     12 others
ORIGIN


  Query Match             1.8%;  Score 73.2;  DB 88;  Length 399;
  Best Local Similarity   57.1%;  Pred. No. 1.1e-10;
  Matches  129;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy   1280 cgcaggctcttagcaaggagggttggcatatatccaagactcttttgacggagataacaa 1339
          | || |  || |||| ||| |      ||    ||||   |    |||||| |    |
Db      7 CCCAAGGCCTAAGCAGCGAGACTGNTNCCATCATGAAGAACATGGCCACGGAGCTGGGCA 66

Qy   1340 ttttgacaattggtcttgctactttcgtgcccgtcatccaggagttttgtatctttgcca 1399
          |  |    || ||       ||  | |||||| ||||||||||  ||| ||| | |
Db     67 TCATCCTCATCGGCTACTTCACCCTAGTGCCCGCCATCCAGGAGTCCTGTCTCTNTCCTG 126

Qy   1400 tagtcggcttgctttccgattttatgctacagatgctgctcttctcaacaatactggcca 1459
          | || ||  || | || || ||  | || |||||||| | ||| | ||  | ||| |||
Db    127 TCGTGGGGCTGGTGTCTGACTTCTTCCTTCAGATGCTGTTTTTCACCACTGTCCTGTCCA 186

Qy   1460 tgaacattaagcggaccgagtatacggcggaggccaagcaccttcc 1505
          |  |||||   |||| |||    | |    |  |||||  || ||
Db    187 TTGACATTCGCCGGATGGAGCTAGCAGACCTGAACAAGCGACTNCC 232


RESULT   6
AA498678
LOCUS       AA498678      561 bp    mRNA            EST       01-JUL-1997
DEFINITION  vh33d06.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA
            clone IMAGE:888779 5' similar to TR:G1228047 G1228047 KIAA0199
            PROTEIN ;, mRNA sequence.
ACCESSION   AA498678
VERSION     AA498678.1  GI:2233701
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 561)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:516739
            Possible reversed clone: similarity on wrong strand
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 377.
FEATURES             Location/Qualifiers
     source          1. .561
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:888779"
                     /clone_lib="Barstead mouse pooled organs MPLRB4"
                     /sex="mixed"
                     /tissue_type="pooled organs"
                     /dev_stage="7 day"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [GTTGGATTCGGTACC], digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library constructed by Bob Barstead."
BASE COUNT      118 a    163 c    154 g    126 t
ORIGIN


  Query Match             1.5%;  Score 61.4;  DB 27;  Length 561;
  Best Local Similarity   50.1%;  Pred. No. 4.3e-07;
  Matches  183;  Conservative    0;  Mismatches  176;  Indels    6;  Gaps    1;

Qy   3109 gcccatcagcagccaatcacatgcatgcaggtcgttaacgacatggttttcactggcagc 3168
          || || ||| |||| |||||| | ||   |  | |  ||       |  | || || |||
Db    118 GCACACCAGAAGCCCATCACAGCCCTGAGAGCTGCTGCCGGGCGCCTAGTGACAGGGAGC 177

Qy   3169 caggatcacaccctcaaggtgtattgcctcaataagtcggatgttgagtatacgctccac 3228
          || || || || || |  || |   | ||  || | ||  |      | ||| ||  |
Db    178 CAAGACCATACTCTAAGAGTCTTCCGACTGGATGACTCGTGTTGCCTCTTTACCCTGAAG 237

Qy   3229 ggtcactgtgggcctgtaacctgtctctttgtggatcgctggcaacctggcacaggggggg 3288
          || |||| ||| |  | ||    | | |   | |||      ||  |      |  ||
Db    238 GGCCACTCAGGGGCAATCACAGCTGTGTACATTGAT------CAGACCATGGTACTGGCC 291

Qy   3289 tctgggtcccaggacggcctgctctgcgtatgggatctgttcacgggagcctgcatgtat 3348
            |||    || || ||     ||||| | |||||| |  | || ||   | |  |
Db    292 AGTGGAGGACAAGATGGAGCCATCTGCCTGTGGGATGTACTAACAGGCAGCCGGGTCAGG 351

Qy   3349 aatatacaagctcacgacggagccgtcagctgcctggcctgtgcgcccagttacgtaatc 3408
            | | |   ||||||  ||||  || | || |||   |||| |  |    |  || ||
Db    352 CAAACATTTGCTCACCGTGGAGATGTTACCTCCCTCAGCTGTACCGCTTCCTGTGTCATT 411

Qy   3409 tcgctaggcacggacgagaggatttgcgtatgggaacgatttcagggaaacctgttgact 3468
               |||  ||| ||    ||  |  | ||||| ||       || | |  | ||
Db    412 AGTAGTGGCCTGGATGACTTCATCAGTATCTGGGAGCGGAGCACAGGCATCAAGCTGTAC 471

Qy   3469 accat 3473
          |||||
Db    472 TCCAT 476


RESULT   7
AA106247
LOCUS       AA106247      432 bp    mRNA            EST       04-FEB-1997
DEFINITION  mm18g04.r1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA
            clone IMAGE:521910 5' similar to TR:G1228047 G1228047 KIAA0199
            PROTEIN ;, mRNA sequence.
ACCESSION   AA106247
VERSION     AA106247.1  GI:1656230
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 432)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:315758
```

```
           Seq primer: -28m13 rev1 ET from Amersham
           High quality sequence stop: 355.
FEATURES             Location/Qualifiers
     source          1. .432
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:521910"
                     /clone_lib="Stratagene mouse diaphragm (#937303)"
                     /tissue_type="diaphragm"
                     /dev_stage="adult"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: diaphragm; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally from mRNA
                     prepared from diaphragm muscle.  Primer: Oligo dT. Average
                     insert size: 1.5 kb. Uni-ZAP XR Vector; ~5' adaptor
                     sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       97 a    121 c    117 g     97 t
ORIGIN


  Query Match             1.5%;  Score 61.2;  DB 21;  Length 432;
  Best Local Similarity   50.0%;  Pred. No. 4.5e-07;
  Matches  184;  Conservative    0;  Mismatches  178;  Indels    6;  Gaps    1;

Qy   3106 ctagcccatcagcagccaatcacatgcatgcaggtcgttaacgacatggttttcactggc 3165
          ||   | | ||| |||| ||||||  | ||   |  | |  ||       |  | || ||
Db     13 CTGTGCAACCAGAAGCCCATCACAGCCCTGAGAGCTGCTGCCGGGCGCCTAGTGACAGGG 72

Qy   3166 agccaggatcacaccctcaaggtgtattgcctcaataagtcggatgttgagtatacgctc 3225
          ||||| || || || || |  || |   | ||  || | |||  |      | ||| ||
Db     73 AGCCAAGACCATACTCTAAGAGTCTTCCGACTGGATGACTCGTGTTGCCTCTTTACCCTG 132

Qy   3226 cacggtcactgtgggcctgtaacctgtctctttgtggatcgctggcaacctggcacaggg 3285
          | || ||||  ||| |  | ||   | | |   | |||      ||  |      |  |
Db    133 AAGGGCCACTCAGGGGCAATCACAGCTGTGTACATTGAT------CAGACCATGGTACTG 186

Qy   3286 gggtctgggtcccaggacggcctgctctgcgtatgggatctgttcacgggagcctgcatg 3345
          |   |||    || || ||    ||||| | ||||||  |  | || ||   | |  |
Db    187 GCCAGTGGAGGACAAGATGGAGCCATCTGCCTGTGGGATGTACTAACAGGCAGCCGGGTC 246

Qy   3346 tataatatacaagctcacgacggagccgtcagctgcctggcctgtgcgcccagttacgta 3405
          | | |   ||||||   ||||  || | || |||  ||||| |  |    |  ||
Db    247 AGCCAAACATTTGCTCACCGTGGAGATGTTACCTCCCTCACCTGTACCGCTTCCTGTGTC 306

Qy   3406 atctcgctaggcacggacgagaggatttgcgtatgggaacgatttcagggaaacctgttg 3465
          ||       |||  ||| ||    ||  |  | ||||| ||       || | |  | ||
Db    307 ATTAGTAGTGGCCTGGATGACTTCATCAGTATCTGGGACCGCAGCACAGGCATCAAGCTG 366

Qy   3466 actaccat 3473
              ||||
Db    367 TACTCCAT 374


RESULT   8
AA254928
LOCUS       AA254928      647 bp    mRNA            EST       14-MAR-1997
DEFINITION  mz84c05.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:720104 5'
            similar to TR:G1228047 G1228047 KIAA0199 PROTEIN ;, mRNA sequence.
ACCESSION   AA254928
VERSION     AA254928.1  GI:1889532
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 647)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:445600
            Possible reversed clone: similarity on wrong strand
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 438.
FEATURES             Location/Qualifiers
     source          1. .647
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:720104"
                     /clone_lib="Soares mouse NML"
                     /tissue_type="Liver"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATCTTTTTTTTTTTTTTTT 3']
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT      137 a    187 c    177 g    145 t      1 others
ORIGIN


  Query Match             1.5%;  Score 60.4;  DB 23;  Length 647;
  Best Local Similarity   49.9%;  Pred. No. 9.2e-07;
  Matches  182;  Conservative    0;  Mismatches  177;  Indels    6;  Gaps    1;

Qy   3109 gcccatcagcagccaatcacatgcatgcaggtcgttaacgacatggttttcactggcagc 3168
          || || ||| |||| ||||||  | ||   |  | |  ||       |  | || || |||
Db    279 GCACACCAGAAGCCCATCACAGCCCTGAGAGCTGCTGCCGGGCGCCTAGTGACAGGGAGC 338

Qy   3169 caggatcacaccctcaaggtgtattgcctcaataagtcggatgttgagtatacgctccac 3228
          || || || || || |  || |   | ||  || | |||  |      | ||| ||  |
Db    339 CAAGACCATACTCTAAGAGTCTTCCGACTGGATGACTCGTGTTGCCTCTTTACCCTGAAG 398

Qy   3229 ggtcactgtgggcctgtaacctgtctctttgtggatcgctggcaacctggcacaggggg 3288
          || ||||  ||| |  | ||   | | |   | |||      ||  |      |  ||
Db    399 GGCCACTCAGGGGCAATCACAGCTGTGTACATTGAT------CAGACCATGGTACTGGCC 452

Qy   3289 tctgggtcccaggacggcctgctctgcgtatgggatctgttcacgggagcctgcatgtat 3348
            |||    || || ||     ||||| | |||||| |  | || ||   | |  |
Db    453 AGTGGAGGACAAGATGGAGCCATCTGCCTGTGGGATGTACTAACAGGCAGCCGGGTCAGC 512

Qy   3349 aatatacaagctcacgacggagccgtcagctgcctggcctgtgcgcccagttacgtaatc 3408
          | | |   ||||||   ||||  || | || |||  ||||| |  |    |  || ||
Db    513 CAAACATTTGCTCACCGTGGAGATGTTACCTCCCTCACCTGTACCGCTTCCTGTGTCATT 572

Qy   3409 tcgctaggcacggacgagaggatttgcgtatgggaacgatttcagggaaacctgttgact 3468
               |    ||| ||    ||  |  | ||||| ||       || | |  | ||
Db    573 AGTAGTGNGCTGGATGACTTCATCAGTATCTGGGACCGCAGCACAGGCATCAAGCTGTAC 632

Qy   3469 accat 3473
           ||||
Db    633 TCCAT 637


RESULT   9
AL043581
LOCUS       AL043581      611 bp    mRNA            EST       29-FEB-2000
DEFINITION  DKFZp434F0527_r1 434 (synonym: htes3) Homo sapiens cDNA clone
            DKFZp434F0527 5', mRNA sequence.
```

```
ACCESSION   AL043581
VERSION     AL043581.1  GI:5422968
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 611)
  AUTHORS   Bloecker,H., Boecher,M., Brandt,P., Mewes,H.W., Gassenhuber,J. and
            Wiemann,S.
  TITLE     EST (Bloecker, et al.)
  JOURNAL   Unpublished (1999)
COMMENT     On May 11, 1999 this sequence version replaced gi:4776630.
            Contact: Bloecker H
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by GBF (National Research Centre for Biotechnology Ltd.,
            Braunschweig/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            s1 sequence also available.
            This clone (DKFZp434F0527) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .611
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp434F0527"
                     /clone_lib="434 (synonym: htes3)"
                     /tissue_type="testis"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT      120 a    178 c    181 g    132 t
ORIGIN
```

```
  Query Match             1.5%;  Score 60.2;  DB 47;  Length 611;
  Best Local Similarity   50.1%;  Pred. No. 1e-06;
  Matches  180;  Conservative    0;  Mismatches  173;  Indels    6;  Gaps    1;

Qy   3115 cagcagccaatcacatgcatgcaggtcgttaacgacatggttttcactggcagccaggat 3174
          ||  | || ||||||  | || | ! || |   |      |  | ||||| ||||| ||
Db      1 CAAAAACCCATCACAGCCCTGAAAGCCGCTGCTGGGCGCTTGGTGACTGGGAGCCAAGAC 60

Qy   3175 cacaccctcaaggtgtattgcctcaataagtcggatgttgagtatacgctccacggtcac 3234
          ||||| || |  |||   | ||  |  | |||        |  || || || || |||
Db     61 CACACACTGAGAGTGTTCCGTCTGGAGGACTCGTGCTGCCTCTTCACCCTTCAGGGCCAC 120

Qy   3235 tgtgggcctgtaacctgtctctttgtggatcgctggcaacctggcacaggggggtctggg 3294
          |  ||| |  | ||     | |   | ||       ||  |          ||    |||
Db    121 TCAGGGGCCATCACGACCGTGTACATTGAC------CAGACCATGGTGCTGGCCAGTGGA 174

Qy   3295 tcccaggacggcctgctctgcgtatgggatctgttcacgggagcctgcatgtataatata 3354
             || || ||     ||||| | |||||| |  | || ||   | |  |     || |
Db    175 GGACAAGATGGGGCCATCTGCCTGTGGGATGTACTGACTGGCAGCCGGGTCAGCCATGTG 234

Qy   3355 caagctcacgacggagccgtcagctgcctggcctgtgcgcccagttacgtaatctcgcta 3414
             ||||||   || |  |||| || |||  ||||| |  ||   |  || |||
Db    235 TTTGCTCACCGTGGGGATGTCACCTCCCTTACCTGTACCACCTCCTGTGTCATCAGCAGT 294

Qy   3415 ggcacggacgagaggatttgcgtatgggaacgatttcagggaaacctgttgactaccat 3473
          |||  ||| ||    ||  || | ||||| ||       || | |  |||     ||||
Db    295 GGCCTGGATGACCTCATCAGCATCTGGGACCGCAGCACAGGCATCAAGTTCTACTCCAT 353
```

```
RESULT  10
AA354156
LOCUS       AA354156      333 bp    mRNA            EST       21-APR-1997
DEFINITION  EST62373 Jurkat T-cells V Homo sapiens cDNA 5' end similar to
            similar to cell division control protein CDC4, mRNA sequence.
ACCESSION   AA354156
VERSION     AA354156.1  GI:2006473
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 333)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., ,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   12140200
COMMENT     On Jan 17, 1998 this sequence version replaced gi:1900162.
            Other_ESTs: THC194741
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .333
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):154386"
                     /db_xref="taxon:9606"
                     /clone_lib="Jurkat T-cells V"
                     /cell_type="T-lymphocyte"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       65 a    105 c     91 g     68 t      4 others
ORIGIN
```

```
  Query Match             1.3%;  Score 51.8;  DB 25;  Length 333;
  Best Local Similarity   49.8%;  Pred. No. 0.00028;
  Matches  156;  Conservative    0;  Mismatches  151;  Indels    6;  Gaps    1;

Qy   3109 gcccatcagcagccaatcacatgcatgcaggtcgttaacgacatggttttcactggcagc 3168
          || || ||  | || ||||||  | || | | || |   |       |  | ||||| |||
Db     23 GCACACCAAAAACCCATCACAGCCCTGAAAGCCGCTGCTGGGCGCTTGGTGACTGGGAGC 82

Qy   3169 caggatcacaccctcaaggtgtattgcctcaataagtcggatgttgagtatacgctccac 3228
          || || ||||| || |  || |   | ||  |  | |||         |  || || ||
Db     83 CAAGACCACACACTGAGAGTNTTCCGTCTGGAGGACTCGTGCTGCCTCTTCACCCTTCAG 142

Qy   3229 ggtcactgtgggcctgtaacctgtctctttgtggatcgctggcaacctggcacaggggggg 3288
          || ||||  ||| |  | ||     | |   | ||       ||  |         ||
Db    143 GGCCACTCAGGGGCCATCACGACCGTNTACATTGAC------CAGACCATGGTGCTGGCC 196

Qy   3289 tctgggtcccaggacggcctgctctgcgtatgggatctgttcacgggagcctgcatgtat 3348
```

```
            |||    || || ||     ||||| | |||||| |  | || ||   | |  |
Db    197 AGTGGAGGACAAGATGGGGCCATCTGCCTGTGGGATGTACTGACTGGCAGCCGGGTCAGC 256

Qy   3349 aatatacaagctcacgacggagccgtcagctgcctggcctgtgcgcccagttacgtaatc 3408
          || |    ||||||   || |  |||| || |||  ||||| |  ||   |   | |||
Db    257 CATGTNTTTGCTCACCGTGGGGATGTCACCTCCCTTACCTGTACCACCTCCTGTNTCATC 316

Qy   3409 tcgctaggcacgg 3421
               |||  ||
Db    317 AGCAGTGGCCTGG 329


RESULT  11
CNS001UZ
LOCUS       CNS001UZ     1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR05H11 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL061599
VERSION     AL061599.1  GI:4939961
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR05H11"
                     /note="end : TET3"
BASE COUNT      374 a    161 c    199 g    307 t     60 others
ORIGIN


  Query Match             1.2%;  Score 50;  DB 122;  Length 1101;
  Best Local Similarity   65.4%;  Pred. No. 0.0015;
  Matches   68;  Conservative    3;  Mismatches   33;  Indels    0;  Gaps    0;

Qy     40 ccaaatagtttttggtatacggatagaatttggatgaaaaataaaacgaaatcaaaacat 99
          ||||     |||| | | |  ||  ||| ||  |  || || ||||  :| | |||| |:
Db    142 CCAACATTCTTTTAGCAGATCGAAGGAAATTTAAAAAACAAAAAAAACWATTTAAAAAAW 201

Qy    100 ttttcaaaagcgtggaagttttggccggcttgtgggcatggcaa 143
          :|||||||||||||| |||||||| ||| |||| ||  |   ||
Db    202 WTTTCAAAAGCGTGGGAGTTTTGGGCGGTTTGTTGGTGTTAWAA 245


RESULT  12
CNS0163J/c
```

```
LOCUS       CNS0163J     1201 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN15O21 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106249
VERSION     AL106249.1  GI:5620937
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN15O21"
                     /note="end : SP6"
BASE COUNT      335 a    278 c     92 g    352 t    144 others
ORIGIN


  Query Match             1.2%;  Score 49.4;  DB 123;  Length 1201;
  Best Local Similarity   67.2%;  Pred. No. 0.0024;
  Matches   84;  Conservative    0;  Mismatches   38;  Indels    3;  Gaps    1;

Qy      8 tgactgttttgtaggtgtaaggaggggcgtggccaaatagtttttggtatacggatagaa 67
          ||   || |||||| | |||||||||||||||| ||||  |||||||||| | |||| ||
Db    432 TGGGCGTATTGTAGCTATAAGGAGGGGCGTGGNCAAAGTGTTTTTGGTATNCCGATAAAA 373

Qy     68 tttggatgaaaaataaaacgaaatcaaaacatttttcaaaagcgtggaagttttggccgg 127
           |||     |||| | ||  |||  || | || |  |||||   |  |||| | |||  |
Db    372 ATTGTC---AAAAAACAATAAAACGAAGAGATGTAACAAAATTTTAAAAGTGTGGGCATG 316

Qy    128 cttgt 132
             ||
Db    315 GCAGT 311


RESULT  13
CNS00AX2/c
LOCUS       CNS00AX2     1101 bp    DNA             GSS       04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR22N02 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL056044
VERSION     AL056044.1  GI:4936914
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
```

```
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            Pl and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .1101
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone_lib="RPCI-98"
                      /clone="BACR22N02"
                      /note="end : T7"
BASE COUNT      303 a    194 c    187 g    350 t     67 others
ORIGIN


  Query Match             1.2%;  Score 48.4;  DB 122;  Length 1101;
  Best Local Similarity   64.3%;  Pred. No. 0.0047;
  Matches  108;  Conservative    2;  Mismatches   43;  Indels   15;  Gaps    2;

Qy       8 tgactgttttgtaggtgtaaggaggggcgtggccaaatagtttttggtatacggatagaa 67
           ||  |||:|| ||| ||||| ||||||||||||   |  |||||||| |||   ||| ||
Db     531 TGKGTGTKTTTTAGCTGTAATGAGGGGCGTGGCACCAGTGTTTTTGGCATAACAATAAAA 472

Qy      68 tttgg----------atgaaaaataaaacgaaatcaaaacatttttcaaaagcgtggaag 117
            || |          |  |||| | |||  ||||| |||| ||||  ||||| ||||  |
Db     471 ATTAGCAAGACAAACAATAAAACTGAAAAAAAATCTAAACGTTTTAAAAAAGTGTGGGTG 412

Qy     118 ttttggccggcttgtgggcatggcaaaacgttttttggctatccgtta 165
           | |  |     || ||||| ||||||||  :|||||  || |  |  ||
Db     411 TATCAG-----TTTTGGGCTTGGCAAAAAKTTTTTAAGCAAATCAATA 369


RESULT  14
CNS00ERH
LOCUS       CNS00ERH     1101 bp    DNA             GSS       04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29C02 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069558
VERSION     AL069558.1  GI:4949701
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            Pl and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .1101
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone_lib="RPCI-98"
                      /clone="BACR29C02"
                      /note="end : T7"
BASE COUNT      364 a    173 c    185 g    281 t     98 others
ORIGIN


  Query Match             1.2%;  Score 48;  DB 122;  Length 1101;
  Best Local Similarity   55.0%;  Pred. No. 0.0061;
  Matches  66;  Conservative    6;  Mismatches   48;  Indels    0;  Gaps    0;

Qy      24 gtaaggaggggcgtggccaaatagtttttggtatacggatagaatttggatgaaaaataa 83
           |   | | | |    ||||  :       | |  ||  ||| |:  |  || || |:
Db     156 GNGNGAANGAGAAANNCCAACAWWCNNNNANNANAWCGAAGGAAATWTAAAAAACAAAAW 215

Qy      84 aacgaaatcaaaacatttttcaaaagcgtggaagttttggccggcttgtgggcatggcaa 143
           ||  :| | |||| ||:||:|||||||||||| |||||||| ||| ||!| ||  |   ||
Db     216 AAACWATTTAAAAAATWTTYCAAAAGCGTGGGAGTTTTGGGCGGTTTGTTGGTGTTATAA 275


RESULT  15
CNS00EQ8
LOCUS       CNS00EQ8      982 bp    DNA             GSS       04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR29P21 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069513
VERSION     AL069513.1  GI:4949656
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 982)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            Pl and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .982
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone_lib="RPCI-98"
                      /clone="BACR29P21"
                      /note="end : TET3"
```

```
BASE COUNT      332 a    138 c    159 g    274 t     79 others
ORIGIN


  Query Match             1.1%;  Score 47.2;  DB 122;  Length 982;
  Best Local Similarity   64.4%;  Pred. No. 0.01;
  Matches   67;  Conservative    2;  Mismatches   35;  Indels    0;  Gaps    0;

Qy      40 ccaaatagtttttggtatacggatagaatttggatgaaaaataaaacgaaatcaaaacat 99
           ||||     |||| | | |  ||  ||| ||  |  || || | ||   | | |||| |:
Db      51 CCAACATTCTTTTAGCAGATCGAAGGAAATTTAAAAAACAAAATAAACTATTTAAAAAAW 110

Qy     100 ttttcaaaagcgtggaagttttggccggcttgtgggcatggcaa 143
           |||||||||:|||||  |||||||| ||| |||| ||  |   ||
Db     111 TTTTCAAAARCGTGGGAGTTTTGGGCGGTTTGTTGGTGTTATAA 154


Search completed: September  4, 2000, 00:25:16
Job time: 19891 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2000, 00:30:21 ; Search time 183.89 Seconds
(without alignments)
3087.881 Million cell updates/sec

Title: US-09-332-522B-5
Perfect score: 4128
Sequence: 1 gtgtgcctgactgttttgta.....tgactaaatgttaaaataaa 4128

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	40.8	1.0	7218	1	US-08-232-463-14	Sequence 14, Appl
c	2	39.4	1.0	7218	1	US-08-232-463-14	Sequence 14, Appl
	3	35.2	0.9	4195	2	US-08-340-011-1	Sequence 1, Appli
	4	35.2	0.9	4425	1	US-08-222-616-31	Sequence 31, Appl
	5	35.2	0.9	4425	6	PCT-US95-04228-31	Sequence 31, Appl
	6	35.2	0.9	4795	2	US-08-340-011-3	Sequence 3, Appli
	7	35.2	0.9	9108	6	PCT-US95-04228-45	Sequence 45, Appl
c	8	34.4	0.8	1201	5	US-09-274-642-1	Sequence 1, Appli
c	9	34.4	0.8	2375	5	US-08-714-918-81	Sequence 81, Appl
	10	34.4	0.8	2447	4	US-09-014-969-14	Sequence 14, Appl
	11	33.6	0.8	790	2	US-08-393-985-22	Sequence 22, Appl
	12	33.6	0.8	2097	2	US-08-393-985-1	Sequence 1, Appli
c	13	33	0.8	10952	2	US-08-602-036A-1	Sequence 1, Appli
c	14	33	0.8	10952	3	US-08-502-374A-1	Sequence 1, Appli
c	15	33	0.8	10952	3	US-08-642-407A-1	Sequence 1, Appli
	16	32.8	0.8	1689	1	US-07-991-867B-41	Sequence 41, Appl
	17	32.8	0.8	1689	4	US-08-544-332-41	Sequence 41, Appl
	18	32.8	0.8	8457	1	US-07-991-867B-1	Sequence 1, Appli
	19	32.8	0.8	8457	4	US-08-544-332-1	Sequence 1, Appli
c	20	32.4	0.8	12124	1	US-08-181-271A-36	Sequence 36, Appl
c	21	32.4	0.8	12124	1	US-08-449-315-36	Sequence 36, Appl
c	22	32.4	0.8	12124	1	US-08-444-803-36	Sequence 36, Appl
c	23	32.4	0.8	12124	1	US-08-449-043-36	Sequence 36, Appl
c	24	32.4	0.8	12124	2	US-08-456-265A-36	Sequence 36, Appl
c	25	32.4	0.8	12124	2	US-08-455-416-36	Sequence 36, Appl
c	26	32.4	0.8	12124	2	US-08-455-244-36	Sequence 36, Appl
c	27	32.4	0.8	12124	2	US-08-454-876-36	Sequence 36, Appl
c	28	32.4	0.8	12124	3	US-08-457-364-36	Sequence 36, Appl
c	29	32.4	0.8	12124	3	US-08-456-262-36	Sequence 36, Appl
c	30	32.4	0.8	12124	3	US-08-456-240-36	Sequence 36, Appl
c	31	32.4	0.8	12124	3	US-08-455-736-36	Sequence 36, Appl
c	32	32.4	0.8	12124	4	US-08-971-217-36	Sequence 36, Appl
	33	32.4	0.8	20303	1	US-08-370-975B-6	Sequence 6, Appli
	34	32.4	0.8	26764	1	US-08-370-975B-1	Sequence 1, Appli
c	35	32.2	0.8	2338	2	US-08-425-069-1	Sequence 1, Appli
c	36	32.2	0.8	2338	4	US-08-317-844B-1	Sequence 1, Appli
	37	32.2	0.8	4008	5	US-08-307-896-5	Sequence 5, Appli
	38	32.2	0.8	4008	6	PCT-US95-11808-5	Sequence 5, Appli
	39	31.6	0.8	1613	3	US-08-813-940-1	Sequence 1, Appli
c	40	31.6	0.8	7225	3	US-08-286-819A-15	Sequence 15, Appl
c	41	31.6	0.8	7225	5	US-08-980-357-15	Sequence 15, Appl
	42	31.4	0.8	500	5	US-09-141-000-2	Sequence 2, Appli
	43	31.4	0.8	2634	5	US-08-911-853-30	Sequence 30, Appl
	44	31.4	0.8	17612	5	US-08-911-853-29	Sequence 29, Appl
	45	31.2	0.8	1165	2	US-08-624-125-11	Sequence 11, Appl

ALIGNMENTS

```
RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-Fls
US-08-232-463-14
```

```
  Query Match             1.0%;  Score 40.8;  DB 1;  Length 7218;
  Best Local Similarity   6.6%;  Pred. No. 0.059;
  Matches   15;  Conservative  128;  Mismatches   85;  Indels    0;  Gaps    0;

Qy     973 tacaggatgtgggagctggtgccttacacagtggcctttatgttggtgtttgcttatgtg 1032
           | ||||  | |||||||| | |  :: : :  :  ::::: : ::  : ::: ::: : :
Db    1044 TGCAGGTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1103

Qy    1033 tacttctctgttcgaaaaatcgatgtatttcgttcccgctttttgctggccttatgtagc 1092
           : :::::: :::      ::  : : :::: ::::: :::::: ::  :::: : :  :
Db    1104 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1163

Qy    1093 gtaatcaccacagccgggagcttggccatgtcccttggcttgtgtttcttctttggcctg 1152
            :  :: :: :  ::     :::  :: : ::::::  ::: : :::::::::::  :::
Db    1164 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1223

Qy    1153 acaatttcgctgcagtcaaaggacattttcccctaccttgtaatcctt 1200
            :  :::: :: :  ::      : :::::::::: :::: :  :::::
Db    1224 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1271


RESULT   2
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             1.0%;  Score 39.4;  DB 1;  Length 7218;
  Best Local Similarity   3.3%;  Pred. No. 0.16;
  Matches   13;  Conservative  210;  Mismatches  166;  Indels    0;  Gaps    0;

Qy    2744 ataatctaatcgcagtaggctgcgccaacggccgcgtagaattgtgggaatcccctgcgg 2803
           | ||| |  |   :: :::  : :  :: ::  : : ::::  : :::::      : ::
Db    1446 AGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1387

Qy    2804 gattgcttaagtgtgcataccaggaagacgcgaagagaaaccagggtataacccacatcc 2863
           ::   :   ::: : : : :  :::::::: : ::::::::::  :::: : ::   : :
Db    1386 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1327

Qy    2864 acctgaacggcgatcgagtgattgtggcgcgtcttaatggccgactagatttttaccgct 2923
           :   ::: :: ::  ::: ::  : :: : :    :: ::  ::  :::     :  :
Db    1326 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1267

Qy    2924 tagagacgtactacaaggggaagcaaatcgactggggttttacctcggcttacaggagaa 2983
            ::::: : :  : ::::::::: :::  ::  ::::   . :    ::   : :::::::
Db    1266 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1207

Qy    2984 ctcatgttcgaactggatccactggaagcctgggattaatgttgcagcagcagcgctgtc 3043
            : :   :::  :::   :  :::::  ::::  :: : :  : :: :: :: :  :
Db    1206 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1147

Qy    3044 agcaagaagcatcccagaagaccaccaaggaggaaatgaaaatcacattggagggtgtaa 3103
           :: :::::: :    ::::::  :  :::::::::: :::::  : :  :::::: : ::
Db    1146 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1087

Qy    3104 gactagcccatcagcagccaatcacatgc 3132
           ::  ::   :  :: ::  :||| || ||
Db    1086 RRRRRRRRRRRRRRRRRRRATCGCAAGC 1058


RESULT   3
US-08-340-011-1
; Sequence 1, Application US/08340011
; Patent No. 5776755
;  GENERAL INFORMATION:
;    APPLICANT:  Alitalo, et al.
;    TITLE OF INVENTION:  FLT4, A NOVEL RECEPTOR TYROSIN KINASE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/340,011
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959,951
;      FILING DATE:  09-OCT-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Thomas C.
;      REGISTRATION NUMBER:  36,989
;      REFERENCE/DOCKET NUMBER:  32267
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4195 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20..3916
US-08-340-011-1


  Query Match             0.9%;  Score 35.2;  DB 2;  Length 4195;
  Best Local Similarity   60.4%;  Pred. No. 2.2;
  Matches   58;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy   1875 acggcgactacaggcgggtcggggagcagtcagcagttttttcgagggatggcaagcgga 1934
           ||   ||| | |||| | | | |||  |||| | |||    | |  |||   ||| | ||
Db   2154 ACAAAGACGAGAGGCTGCTGGAGGAAAAGTCTGGAGTCGACTTGGCGGACTCCAACCAGA 2213

Qy   1935 cgggcagcgtgccacgagtgcgccaagcggaagcgg 1970
           |   ||| | |  || |||||| | | ||| ||||
Db   2214 AGCTGAGCATCCAGCGCGTGCGCGAGGAGGATGCGG 2249


RESULT   4
US-08-222-616-31
; Sequence 31, Application US/08222616
; Patent No. 5635177
;  GENERAL INFORMATION:
;    APPLICANT:  Bennett, Brian D.
;    APPLICANT:  Goeddel, David
;    APPLICANT:  Lee, James M.
;    APPLICANT:  Matthews, William
;    APPLICANT:  Tsai, Siao Ping
;    APPLICANT:  Wood, William I.
;    TITLE OF INVENTION:  PROTEIN TYROSINE KINASE AGONIST
;    TITLE OF INVENTION:  ANTIBODIES
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/222,616
;      FILING DATE:  4-APR-1994
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/00586
;      FILING DATE:  22-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/826935
;      FILING DATE:  22-JAN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lee, Wendy M.
;      REGISTRATION NUMBER:
;      REFERENCE/DOCKET NUMBER:  821P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1994
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4425 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-222-616-31


  Query Match             0.9%;  Score 35.2;  DB 1;  Length 4425;
  Best Local Similarity   60.4%;  Pred. No. 2.3;
  Matches   58;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy   1875 acggcgactacaggcgggtcggggagcagtcagcagttttttcgagggatggcaagcgga 1934
           ||   ||| | |||| | | | |||  |||| | |||    | |  |||   ||| | ||
Db   2165 ACAAAGACGAGAGGCTGCTGGAGGAAAAGTCTGGAGTCGACTTGGCGGACTCCAACCAGA 2224

Qy   1935 cgggcagcgtgccacgagtgcgccaagcggaagcgg 1970
           |   ||| | |  || |||||| | | ||| ||||
Db   2225 AGCTGAGCATCCAGCGCGTGCGCGAGGAGGATGCGG 2260


RESULT   5
PCT-US95-04228-31
; Sequence 31, Application PC/TUS9504228
;  GENERAL INFORMATION:
;    APPLICANT:  Genentech, Inc.
;    APPLICANT:  Bennett, Brian D.
;    APPLICANT:  Goeddel, David
;    APPLICANT:  Lee, James M.
;    APPLICANT:  Matthews, William
;    APPLICANT:  Tsai, Siao Ping
;    APPLICANT:  Wood, William I.
;    TITLE OF INVENTION:  PROTEIN TYROSINE KINASE AGONIST ANTIBODIES
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/04228
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/222616
;      FILING DATE:  04-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wendy M. Lee
;      REGISTRATION NUMBER:  00,000
;      REFERENCE/DOCKET NUMBER:  821P3PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1994
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4425 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
PCT-US95-04228-31


  Query Match             0.9%;  Score 35.2;  DB 6;  Length 4425;
  Best Local Similarity   60.4%;  Pred. No. 2.3;
  Matches   58;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy   1875 acggcgactacaggcgggtcggggagcagtcagcagttttttcgagggatggcaagcgga 1934
           ||   ||| | |||| | | | |||  |||| | |||    | |  |||   ||| | ||
Db   2165 ACAAAGACGAGAGGCTGCTGGAGGAAAAGTCTGGAGTCGACTTGGCGGACTCCAACCAGA 2224

Qy   1935 cgggcagcgtgccacgagtgcgccaagcggaagcgg 1970
           |   ||| | |  || |||||| | | ||| ||||
Db   2225 AGCTGAGCATCCAGCGCGTGCGCGAGGAGGATGCGG 2260
```

```
RESULT   6
US-08-340-011-3
; Sequence 3, Application US/08340011
; Patent No. 5776755
;  GENERAL INFORMATION:
;    APPLICANT:  Alitalo, et al.
;    TITLE OF INVENTION:  FLT4, A NOVEL RECEPTOR TYROSIN KINASE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/340,011
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959,951
;      FILING DATE:  09-OCT-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Thomas C.
;      REGISTRATION NUMBER:  36,989
;      REFERENCE/DOCKET NUMBER:  32267
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4795 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20..4111
US-08-340-011-3


  Query Match             0.9%;  Score 35.2;  DB 2;  Length 4795;
  Best Local Similarity   60.4%;  Pred. No. 2.4;
  Matches   58;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy   1875 acggcgactacaggcgggtcggggagcagtcagcagttttttcgagggatggcaagcgga 1934
           ||   ||| | |||| | | | |||  |||| | |||    | |  |||   ||| | ||
Db   2154 ACAAAGACGAGAGGCTGCTGGAGGAAAAGTCTGGAGTCGACTTGGCGGACTCCAACCAGA 2213

Qy   1935 cgggcagcgtgccacgagtgcgccaagcggaagcgg 1970
            |   ||| | |  || |||||| | | ||| ||||
Db   2214 AGCTGAGCATCCAGCGCGTGCGCGAGGAGGATGCGG 2249


RESULT   7
PCT-US95-04228-45
; Sequence 45, Application PC/TUS9504228
;  GENERAL INFORMATION:
;    APPLICANT:  Genentech, Inc.
;    APPLICANT:  Bennett, Brian D.
;    APPLICANT:  Goeddel, David
;    APPLICANT:  Lee, James M.
;    APPLICANT:  Matthews, William
;    APPLICANT:  Tsai, Siao Ping
;    APPLICANT:  Wood, William I.
;    TITLE OF INVENTION:  PROTEIN TYROSINE KINASE AGONIST ANTIBODIES
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/04228
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/222616
;      FILING DATE:  04-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wendy M. Lee
;      REGISTRATION NUMBER:  00,000
;      REFERENCE/DOCKET NUMBER:  821P3PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1994
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9108 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
PCT-US95-04228-45


  Query Match             0.9%;  Score 35.2;  DB 6;  Length 9108;
  Best Local Similarity   60.4%;  Pred. No. 3.6;
  Matches   58;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy   1875 acggcgactacaggcgggtcggggagcagtcagcagttttttcgagggatggcaagcgga 1934
           ||   ||| | |||| | | | |||  |||| | |||    | |  |||   ||| | ||
Db   3097 ACAAAGACGAGAGGCTGCTGGAGGAAAAGTCTGGAGTCGACTTGGCGGACTCCAACCAGA 3156

Qy   1935 cgggcagcgtgccacgagtgcgccaagcggaagcgg 1970
            |   ||| | |  || |||||| | | ||| ||||
Db   3157 AGCTGAGCATCCAGCGCGTGCGCGAGGAGGATGCGG 3192


RESULT   8
US-09-274-642-1/c
; Sequence 1, Application US/09274642A
; Patent No. 6071729
; GENERAL INFORMATION:
;  APPLICANT: Jeffries, Thomas W.
;  APPLICANT: Shi, Nian-Qing
;  TITLE OF INVENTION: Disruption of cytochrome c gene in xylose-fermenting yeast
;  FILE REFERENCE: cytochrome c
;  CURRENT APPLICATION NUMBER: US/09/274,642A
;  CURRENT FILING DATE: 1999-03-23
;  EARLIER APPLICATION NUMBER: 60/080,493
;  EARLIER FILING DATE: 1998-04-02
;  NUMBER OF SEQ ID NOS: 42
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 1201
;   TYPE: DNA
;   ORGANISM: Pichia stipitis
;   FEATURE:
;   NAME/KEY: CDS
```

```
;   LOCATION: (610)..(942)
US-09-274-642-1


  Query Match             0.8%;  Score 34.4;  DB 5;  Length 1201;
  Best Local Similarity   59.0%;  Pred. No. 1.8;
  Matches   59;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy   3535 aaaatgggtaaggcctcattcttgattgccaatataagagggacagtaaataataaattt 3594
          ||| | || ||  |||| | |||   |  ||   | | |    |||  ||||||    |
Db    543 AAATTCGGCAACTCCTCTTCCTTTTATATCAGACTGAAAAATTCAGCCAATAATGCCATA 484

Qy   3595 aattccaacacaggatctcttattgtgtgggatgtgcgca 3634
          | | | |||  | || |||| |||| ||| |||| |||
Db    483 ATTGCTAACCAATGAGCTCTCATTGGATGGAGTGTGTGCA 444


RESULT   9
US-08-714-918-81/c
; Sequence 81, Application US/08714918
; Patent No. 6037123
;  GENERAL INFORMATION:
;    APPLICANT:  Benton, Bret
;    APPLICANT:  Lee, Ving
;    APPLICANT:  Malouin, Francois
;    APPLICANT:  Martin, Patrick K.
;    APPLICANT:  Schmid, Molly B.
;    APPLICANT:  Sun, Dongxu
;    TITLE OF INVENTION:  STAPHYLOCOCCUS AUREUS ANTIBACTERIAL
;    TITLE OF INVENTION:  TARGET GENES
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/714,918
;      FILING DATE:  September 13, 1996
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/009,102
;      FILING DATE:  December 22, 1995
;      APPLICATION NUMBER:  60/003,798
;      FILING DATE:  September 15, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  222/005
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  81:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2375 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-714-918-81


  Query Match             0.8%;  Score 34.4;  DB 5;  Length 2375;
  Best Local Similarity   60.9%;  Pred. No. 2.8;
  Matches   56;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy   1986 aaaagctcctctagcgatcgatataaacgaaacggccgaggaaatgatgagacttcgata 2045
          ||||||  |  |  |  ||  |  ||  |  ||  |   ||  |  |  |  |||  |||  ||  |  |   |
Db   1556 AAAAGCCCTTGAATCGTGCCATCTTCACCATTAGGACCATGTAATAATGGGAATACTGCA 1497

Qy   2046 tcccagcttcgacctaaactatttcctttcaa 2077
          ||  |   |  |  |||  |||||  ||  |||||||
Db   1496 TCGTATGGTTGTCCTGAACTACTTTCTTTCAA 1465


RESULT  10
US-09-014-969-14
; Sequence 14, Application US/09014969
; Patent No. 5965397
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs, Kenneth
;    APPLICANT:  McCoy, John M.
;    APPLICANT:  LaVallie, Edward R.
;    APPLICANT:  Racie, Lisa A.
;    APPLICANT:  Merberg, David
;    APPLICANT:  Treacy, Maurice
;    APPLICANT:  Spaulding, Vikki
;    APPLICANT:  Agostino, Michael J.
;    TITLE OF INVENTION:  SECRETED PROTEINS AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  ENCODING THEM
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  MA
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/014,969
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sprunger, Suzanne A.
;      REGISTRATION NUMBER:  41,323
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8284
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2447 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-014-969-14


  Query Match             0.8%;  Score 34.4;  DB 4;  Length 2447;
  Best Local Similarity   48.4%;  Pred. No. 2.8;
  Matches   92;  Conservative    1;  Mismatches   97;  Indels    0;  Gaps    0;

Qy   3939 tgtacgtatctcatctctagaattttgtgttagtttgtggacactaagtgtaacagctac 3998
          | || |  || ||| ||| ||  | ||| || | || || ||  ||| ||| |||  ||
Db   2173 TATATGGTTCACATATCTGGATCTGTGTATTTGATTTTGTACTTTAAATGTGACAAATAA 2232

Qy   3999 gctccggtaggttaaggaactaaactaaatgaatcagatatatacacatatattttcgcg 4058
            |      |   ||  ||  |||  :||  ||  | | | | | | | | |
Db   2233 ACCTTTTGGGAGAAAAAAAAAAAAAAARAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2292

Qy   4059 taattatataaactacatagtgtcttaaagcgcctcagcctaatataaaatgactaaatg 4118
           ||  | | |||  | | |       |||       |    || | ||||  |  |||
```

```
Db    2293 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2352

Qy    4119 ttaaaataaa 4128
              |||| |||
Db    2353 AAAAAAAAAA 2362


RESULT 11
US-08-393-985-22
; Sequence 22, Application US/08393985
; Patent No. 5693476
;  GENERAL INFORMATION:
;    APPLICANT:  Scheller, Richard H.
;    TITLE OF INVENTION:  Methods and Compositions for Modulation
;    TITLE OF INVENTION:  of Vesicular Release
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,985
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sholtz, Charles K.
;      REGISTRATION NUMBER:  38,615
;      REFERENCE/DOCKET NUMBER:  8600-0152
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  790 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  Cytoplasmic domain of Rat syntaxin 1A
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..790
US-08-393-985-22


  Query Match             0.8%;  Score 33.6;  DB 2;  Length 790;
  Best Local Similarity   50.6%;  Pred. No. 2.5;
  Matches   81;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy     564 agacaacgtaaagcgcggaacacacgggcagctggaccagatctttccggagtatggctg 623
             || |||  |   |||   |  |     |    | || |||| | |    | || | |    |
Db     247 AGCGAACAAAGTTCGCTCCAAGCTAAAGAGCATCGAGCAGAGCATCGAGCAGGAGGAAGG 306

Qy     624 cctgctgctctcgcccgccaacctgtggacgcagaactctcagaactttactcgggacac 683
              |||    | |||| | ||  ||||| |||  | |||   | ||| | |   || | |  |
Db     307 TCTGAACCGCTCGTCGGCGGACCTGAGGATCCGGAAGACGCAGCATTCCACGCTGTCCCG 366

Qy     684 aaacatcctgaacacgatatttcagtaccataacctacag 723
             |||  |  || |     || |   ||||| |    |   |||
Db     367 AAAGTTTGTGGAGGTCATGTCCGAGTACAACGCCACTCAG 406


RESULT 12
US-08-393-985-1
; Sequence 1, Application US/08393985
; Patent No. 5693476
;  GENERAL INFORMATION:
;    APPLICANT:  Scheller, Richard H.
;    TITLE OF INVENTION:  Methods and Compositions for Modulation
;    TITLE OF INVENTION:  of Vesicular Release
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,985
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sholtz, Charles K.
;      REGISTRATION NUMBER:  38,615
;      REFERENCE/DOCKET NUMBER:  8600-0152
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2097 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  Rat syntaxin 1A 3' end (encoding amino
;      INDIVIDUAL ISOLATE:  acids 4-288; GenBank M95734)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..859
US-08-393-985-1


  Query Match             0.8%;  Score 33.6;  DB 2;  Length 2097;
  Best Local Similarity   50.6%;  Pred. No. 4.5;
  Matches   81;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy     564 agacaacgtaaagcgcggaacacacgggcagctggaccagatctttccggagtatggctg 623
             || |||  |   |||   |  |     |    | || |||| | |    | || | |    |
Db     247 AGCGAACAAAGTTCGCTCCAAGCTAAAGAGCATCGAGCAGAGCATCGAGCAGGAGGAAGG 306

Qy     624 cctgctgctctcgcccgccaacctgtggacgcagaactctcagaactttactcgggacac 683
              |||    | |||| | ||  ||||| |||  | |||   | ||| | |   || | |  |
Db     307 TCTGAACCGCTCGTCGGCGGACCTGAGGATCCGGAAGACGCAGCATTCCACGCTGTCCCG 366

Qy     684 aaacatcctgaacacgatatttcagtaccataacctacag 723
             |||  |  || |     || |   ||||| |    |   |||
Db     367 AAAGTTTGTGGAGGTCATGTCCGAGTACAACGCCACTCAG 406


RESULT 13
US-08-602-036A-1/c
; Sequence 1, Application US/08602036A
; Patent No. 5789248
;  GENERAL INFORMATION:
;    APPLICANT:  Oeystein, Fodstad
```

```
;    APPLICANT:  Hovig, Eivind
;    APPLICANT:  Engebraaten, Olav
;    APPLICANT:  Maelandsmo, Gunhild H.
;    APPLICANT:  Agrawal, Sudhir
;    TITLE OF INVENTION:  CAPL SPECIFIC OLIGONUCLEOTIDES AND
;    TITLE OF INVENTION:  METHODS OF INHIBITING METASTATIC CANCER
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HALE AND DORR LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  United States of America
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/602,036A
;      FILING DATE:  16-FEB-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kerner, Ann-Louise
;      REGISTRATION NUMBER:  33,523
;      REFERENCE/DOCKET NUMBER:  HYZ-039CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 526-6000
;      TELEFAX:  (617) 526-5000
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10952 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-602-036A-1


  Query Match             0.8%;  Score 33;  DB 2;  Length 10952;
  Best Local Similarity   53.5%;  Pred. No. 19;
  Matches   69;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy    881 aaaagctgctgcgacactatcccccactcccgttggctagtgcgtcggctgaagagccga 940
          || | | ||       | | ||| ||||| |  | | | | | | ||| |  | | |
Db   5663 AACAACAGCATTTTTTCCACCCCTCACTCAGTCTATCCACTCCCACAGCTCAGCCACAAA 5604

Qy    941 cgaccataacttacatcttttatccaggagagtacaggatgtgggagctggtgccttaca 1000
          | ||    ||| |   | || |||     || || || |||||  ||||| | | | |
Db   5603 GGCCCCGCCCTTGCCATTCTTTTCCCTCCCAGAACTGGGTGTGGCTGCTGGAGGCCTGCC 5544

Qy   1001 cagtggcct 1009
          ||| |||||
Db   5543 CAGGGGCCT 5535


RESULT  14
US-08-502-374A-1/c
; Sequence 1, Application US/08502374A
; Patent No. 5872007
;  GENERAL INFORMATION:
;    APPLICANT:  Fodstad, Oeystein
;    APPLICANT:  Hovig, Eivind
;    APPLICANT:  Engebraaten, Olav
;    APPLICANT:  Maelandsmo, Gunhild H.
;    APPLICANT:  Agrawal, Sudhir
;    TITLE OF INVENTION:  CAPL-SPECIFIC OLIGONUCLEOTIDES AND
;    TITLE OF INVENTION:  METHODS OF INHIBITING METASTATIC CANCER
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hale and Dorr LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/502,374A
;      FILING DATE:  14-Jul-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kerner, Ann-Louise
;      REGISTRATION NUMBER:  33,523
;      REFERENCE/DOCKET NUMBER:  HYZ-039DV2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-526-6000
;      TELEFAX:  617-526-5000
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10952 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-502-374A-1


  Query Match             0.8%;  Score 33;  DB 3;  Length 10952;
  Best Local Similarity   53.5%;  Pred. No. 19;
  Matches   69;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy    881 aaaagctgctgcgacactatcccccactcccgttggctagtgcgtcggctgaagagccga 940
          || | | ||       | | ||| ||||| |  | | | | | | ||| |  | | |
Db   5663 AACAACAGCATTTTTTCCACCCCTCACTCAGTCTATCCACTCCCACAGCTCAGCCACAAA 5604

Qy    941 cgaccataacttacatcttttatccaggagagtacaggatgtgggagctggtgccttaca 1000
          | ||    ||| |   | || |||     || || || |||||  ||||| | | | |
Db   5603 GGCCCCGCCCTTGCCATTCTTTTCCCTCCCAGAACTGGGTGTGGCTGCTGGAGGCCTGCC 5544

Qy   1001 cagtggcct 1009
          ||| |||||
Db   5543 CAGGGGCCT 5535


RESULT  15
US-08-642-407A-1/c
; Sequence 1, Application US/08642407A
; Patent No. 5877308
;  GENERAL INFORMATION:
;    APPLICANT:  Oeystein, Fodstad
;    APPLICANT:  Hovig, Eivind
;    APPLICANT:  Engebraaten, Olav
;    APPLICANT:  Maelandsmo, Gunhild H.
;    APPLICANT:  Agrawal, Sudhir
;    TITLE OF INVENTION:  CAPL SPECIFIC OLIGONUCLEOTIDES AND
;    TITLE OF INVENTION:  METHODS OF INHIBITING METASTATIC CANCER
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HALE AND DORR LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  United States of America
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/642,407A
;       FILING DATE:  03-May-1996
;       CLASSIFICATION:  514
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kerner, Ann-Louise
;       REGISTRATION NUMBER:  33,523
;       REFERENCE/DOCKET NUMBER:  HYZ-039CPDV
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 526-6000
;       TELEFAX:  (617) 526-5000
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  10952 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  genomic DNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
US-08-642-407A-1


  Query Match             0.8%;  Score 33;  DB 3;  Length 10952;
  Best Local Similarity   53.5%;  Pred. No. 19;
  Matches   69;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy     881 aaaagctgctgcgacactatcccccactcccgttggctagtgcgtcggctgaagagccga 940
           || | | ||       | | ||| ||||     |  | | | |  | ||| |    |  |
Db    5663 AACAACAGCATTTTTTCCACCCCTCACTCAGTCTATCCACTCCCACAGCTCAGCCACAAA 5604

Qy     941 cgaccataacttacatcttttatccaggagagtacaggatgtgggagctggtgccttaca 1000
            | ||    ||| |   | || |||     || || || |||||  ||||| | | | |
Db    5603 GGCCCCGCCCTTGCCATTCTTTTCCCTCCCAGAACTGGGTGTGGCTGCTGGAGGCCTGCC 5544

Qy    1001 cagtggcct 1009
           ||| |||||
Db    5543 CAGGGGCCT 5535


Search completed: September  4, 2000, 00:33:20
Job time: 20170 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2000, 00:39:04 ; Search time 299.13 Seconds
(without alignments)
3452.656 Million cell updates/sec

Title: US-09-332-522B-5
Perfect score: 4128
Sequence: 1 gtgtgcctgactgttttgta.....tgactaaatgttaaaataaa 4128

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	43.4	1.1	2481	1	X28627	Nucleotide sequenc
	2	36.6	0.9	110000	1	V21209_14	Continuation (15 o
	3	35.2	0.9	4195	1	T12068	FLT4 receptor tyro
	4	35.2	0.9	4425	1	T03090	Protein tyrosine-k
	5	35.2	0.9	9108	1	T03104	Plasmid pRK5.tk1-1
c	6	35	0.8	110000	1	V21209_07	Continuation (8 of
c	7	34.4	0.8	279	1	V86487	EST clone AT111. N
	8	34.4	0.8	2447	1	V54587	Human secretory pr
c	9	34.4	0.8	7363	1	V74330	Staphylococcus aur
c	10	34.4	0.8	237326	1	V57903	Hereditary haemoch
c	11	34.2	0.8	376	1	Q39893	Expressed Sequence
c	12	34.2	0.8	376	1	Q59305	Human brain Expres
c	13	34	0.8	3060	1	X07378	Arabidopsis TTG1 g
c	14	34	0.8	3060	1	X07374	Arabidopsis TTG1 g
c	15	34	0.8	3060	1	X07376	Arabidopsis TTG1 g
c	16	34	0.8	3060	1	X07377	Arabidopsis TTG1 g
c	17	34	0.8	3060	1	X07379	Arabidopsis TTG1 g
c	18	34	0.8	3060	1	X07380	Arabidopsis TTG1 g
c	19	34	0.8	3060	1	X07383	Arabidopsis TTG1 g
	20	34	0.8	28690	1	X13075	Enterococcus faeca
	21	33.6	0.8	2097	1	V01546	Rat syntaxin 1A ge
	22	33.6	0.8	7019	1	X13064	Enterococcus faeca
c	23	33.4	0.8	13104	1	Q46852	Clone of recombina
c	24	33	0.8	2522	1	V63196	cDNA from clone fp
	25	33	0.8	2809	1	X13371	Enterococcus faeca
	26	33	0.8	4239	1	V74644	Staphylococcus aur
c	27	33	0.8	10952	1	T33345	Human CAPL gene. S
c	28	33	0.8	10952	1	V41162	Human CAPL genee.
c	29	33	0.8	10952	1	X17696	Human CAPL genomic
	30	32.8	0.8	1689	1	Q66824	AmEPV NPH-1 gene.
c	31	32.8	0.8	2118	1	V84592	Human secreted pro
	32	32.8	0.8	8457	1	Q66797	AmEPV spheroidin g
	33	32.6	0.8	507	1	V24841	H. pylori ORF hp4p
c	34	32.6	0.8	753	1	Q78153	Barley mild mosaic
	35	32.6	0.8	1320	1	V25064	H. pylori cytoplas
c	36	32.6	0.8	3485	1	Q53461	BaMMV coat protein
	37	32.6	0.8	7965	1	V74632	Staphylococcus aur
c	38	32.4	0.8	12124	1	V81625	Cucumber chitinase
c	39	32.4	0.8	12124	1	V62854	Cucumber chitinase
	40	32.4	0.8	20303	1	T71699	Human deoxycytidyl
	41	32.4	0.8	26764	1	T71696	Human deoxycytidyl
	42	32.2	0.8	409	1	V88259	EST clone DY516. N
c	43	32.2	0.8	2338	1	Q14183	N.clavipes draglin
c	44	32.2	0.8	2338	1	V23249	Nephila clavipes s
	45	32.2	0.8	4008	1	T14528	Rat adenylyl cycla

ALIGNMENTS

```
RESULT   1
X28627
ID   X28627 standard; cDNA; 2481 BP.
AC   X28627;
DT   07-JUN-1999  (first entry)
DE   Nucleotide sequence of Sel-10 gene.
KW   Nematode; SEL-10 gene; Alzheimer's disease; cancer; ss.
OS   Caenorhabditis elegans.
FH   Key             Location/Qualifiers
FT   CDS             91. .1854
FT                   /*tag= a
FT                   /product= "Sel-10"
FT   misc_feature    85. .86
FT                   /*tag= b
FT                   /note= "SL1 splice junction"
FT   misc_feature    179. .180
FT                   /*tag= c
FT                   /note= "SL1 splice junction"
FT   misc_feature    247. .248
FT                   /*tag= d
FT                   /note= "SL1 splice junction"
FT   misc_feature    344. .345
FT                   /*tag= e
FT                   /note= "SL1 splice junction"
FT   misc_feature    469. .594
FT                   /*tag= f
FT                   /note= "F-Box"
FT   misc_feature    479. .480
FT                   /*tag= g
FT                   /note= "SL1 splice junction"
FT   misc_feature    625. .626
FT                   /*tag= h
FT                   /note= "SL1 splice junction"
FT   misc_feature    746. .747
FT                   /*tag= i
FT                   /note= "SL1 splice junction"
FT   repeat_region   847. .936
FT                   /*tag= j
FT                   /note= "WD40 (I)"
FT   repeat_region   969. .1062
FT                   /*tag= k
FT                   /note= "WD40 (II)"
FT   misc_difference 1059
FT                   /*tag= l
FT                   /note= "optionally a G to A transition"
FT   repeat_region   1095. .1182
FT                   /*tag= m
FT                   /note= "WD40 (III)"
FT   misc_feature    1136. .1137
FT                   /*tag= n
FT                   /note= "SL1 plice junction"
FT   repeat_region   1215. .1302
FT                   /*tag= o
FT                   /note= "WD40 (IV)"
FT   repeat_region   1335. .1428
FT                   /*tag= p
```

```
FT                   /note= "WD40 (V)"
FT   misc_feature    1362. .1363
FT                   /*tag=  q
FT                   /note= "SL1 splice junction"
FT   repeat_region   1471. .1557
FT                   /*tag=  r
FT                   /note= "WD40 (VI)"
FT   misc_feature    1547. .1548
FT                   /*tag=  s
FT                   /note= "SL1 splice junction"
FT   repeat_region   1590. .1680
FT                   /*tag=  t
FT                   /note= "WD40 (VII)"
FT   misc_difference 1623
FT                   /*tag=  u
FT                   /note= "Optionally a G to A transition"
FT   misc_feature    1688. .1689
FT                   /*tag=  v
FT                   /note= "SL1 splice junction"
FT   polyA_signal    2437. .2442
FT                   /*tag=  w
PN   WO9905307-A1.
PD   04-FEB-1999.
PF   23-JUL-1998; U15335.
PR   24-JUL-1997; US-899578.
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
PI   Greenwald I, Hubbard EJ;
DR   WPI; 99-142956/12.
DR   P-PSDB; Y03204.
PT   New isolated Caenorhabditis elegans SEL-10 gene - used to develop
PT   products which can be used for ameliorating Alzheimer's disease or
PT   for treating cancers
PS   Claim 1; Fig 2A-2C; 85pp; English.
CC   This is the nucleotide sequence of the Caenorhabditis elegans SEL-10
CC   gene. It is used in the method of the invention to develop products
CC   and methods that can be used for identifying compounds which can
CC   ameliorate Alzheimer's disease or which can treat cancer.
SQ   Sequence   2481 BP;    662 A;    517 C;    500 G;    802 T;


  Query Match             1.1%;  Score 43.4;  DB 1;  Length 2481;
  Best Local Similarity   52.5%;  Pred. No. 0.0041;
  Matches   95;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy   3059 agaagaccaccaaggaggaaatgaaaatcacattggagggtgtaagactagcccatcagc 3118
         | |||| |   || | | |    |  ||    |  |  |   || ||  |  | |  | |
Db    800 AAAAGAACTGGAATGCGAATCCAATTATGGGGTCAGCAGTGCTACGAGGACACGAAGATC 859

Qy   3119 agccaatcacatgcatgcaggtcgttaacgacatggttttcactggcagccaggatcaca 3178
          |   |||||| || |||||  |   | | ||  |  |  | |||||     | ||| |||
Db    860 ATGTAATCACTTGTATGCAAATTCATGATGATGTCTTGGTGACTGGATCTGACGATAACA 919

Qy   3179 ccctcaaggtgtattgcctcaataagtcggatgttgagtatacgctccacggtcactgtg 3238
          | || || || |  ||  |  | ||    || |||  ||| || ||   ||| |||  ||
Db    920 CTCTTAAAGTATGGTGTATTGACAAAGGAGAGGTTATGTACACACTAGTCGGCCACACTG 979

Qy   3239 g 3239
          |
Db    980 G 980


RESULT   2
V21209_14
Continuation (15 of 17) of V21209 from base 1400001 (Methanococcus jannaschii circular c
WP Sequence split into 17 fragments  LOCUS V21209 Accession V21209
WP     Fragment Name      Begin     End
WP     V21209_00              1    110000
WP     V21209_01         100001    210000
WP     V21209_02         200001    310000
WP     V21209_03         300001    410000
WP     V21209_04         400001    510000
WP     V21209_05         500001    610000
WP     V21209_06         600001    710000
WP     V21209_07         700001    810000
WP     V21209_08         800001    910000
WP     V21209_09         900001   1010000
WP     V21209_10        1000001   1110000
WP     V21209_11        1100001   1210000
WP     V21209_12        1200001   1310000
WP     V21209_13        1300001   1410000
WP     V21209_14        1400001   1510000
WP     V21209_15        1500001   1610000
WP     V21209_16        1600001   1664976


  Query Match             0.9%;  Score 36.6;  DB 1;  Length 110000;
  Best Local Similarity   53.1%;  Pred. No. 9.7;
  Matches   78;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy   3509 taacaccgtcactattggttacgagcaaaatgggtaaggcctcattcttgattgccaata 3568
         ||| ||| ||| ||   |  | | |  || | | ||  ||  | ||  |    || |  |
Db  51379 TAAGACCTTCATTAAGTGAAATGGGAGAAGTTGATAGAGCTGCTTTAATAGAGGCAAGGA 51438

Qy   3569 taagagggacagtaaataataaatttaattccaacacaggatctcttattgtgtgggatg 3628
           | |   || | | | | || ||||| ||    |||||  || | |  ||| || |||||
Db  51439 AAGGGAAGAAATTCATTTATCAATTTTATAATGACACAACATGTTTGGTTGAGTTGGATG 51498

Qy   3629 tgcgcactgggcagccggctcgcgagg 3655
                ||   || ||  |  | || |
Db  51499 AAGAACCTCCACATCCACCAAGTGAAG 51525


RESULT   3
T12068
ID   T12068 standard; DNA; 4195 BP.
AC   T12068;
DT   08-APR-1996  (first entry)
DE   FLT4 receptor tyrosine kinase DNA.
KW   FLT4; receptor tyrosine kinase; probe; metastasis; lymphoma;
KW   lymphangioma; immunological disease; cancer; diagnosis; therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             20. .3976
FT                   /*tag=  a
PN   WO9533772-A1.
PD   14-DEC-1995.
PF   09-JUN-1995; FI0337.
PR   09-JUN-1994; US-257754.
PA   (ALIT/) ALITALO K.
PA   (KAIP/) KAIPAINEN A.
PA   (KARN/) KARNANI P.
PA   (KORH/) KORHONEN J.
PA   (MATI/) MATIKAINEN M.
PA   (MUST/) MUSTONEN T.
PA   (PAJU/) PAJUSOLA K.
PI   Alitalo K,  Kaipainen A,  Karnani P,  Korhonen J,  Matikainen M;
PI   Mustonen T, Pajusola K;
DR   WPI; 96-040189/04.
DR   P-PSDB; R90528.
PT   Antibodies against FLT4 receptor tyrosine kinase, hybridomas and
PT   nucleic acid probes - used to diagnose and treat e.g. metastatic
PT   cancers, involving alterations to lymphatic vessels.
PS   Claim 9; Page 27-33; 54pp; English.
CC   A DNA sequence (T12068) coding for FLT4 receptor tyrosine kinase
CC   (R90528) is used as a probe that specifically binds/hybridizes to
CC   DNA encoding human FLT4.  Such probes are used to detect FLT4, a
CC   novel marker for lymphatic vessels and some high endothelial
CC   venules, in biological tissue for use in diagnosis and therapy of
CC   e.g. inflammatory, infectious and immunological diseases, metastatic
CC   lymph nodes and lymphangiomas.
SQ   Sequence   4195 BP;    889 A;    1278 C;    1305 G;    723 T;


  Query Match             0.9%;  Score 35.2;  DB 1;  Length 4195;
  Best Local Similarity   60.4%;  Pred. No. 2.6;
  Matches   58;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;
```

```
Qy   1875 acggcgactacaggcgggtcggggagcagtcagcagttttttcgagggatggcaagcgga 1934
          ||   ||| | |||| | | | |||  |||| | |||    | |  |||   ||| | ||
Db   2154 ACAAAGACGAGAGGCTGCTGGAGGAAAAGTCTGGAGTCGACTTGGCGGACTCCAACCAGA 2213

Qy   1935 cgggcagcgtgccacgagtgcgccaagcggaagcgg 1970
           |   ||| | |  || |||||| | | ||| ||||
Db   2214 AGCTGAGCATCCAGCGCGTGCGCGAGGAGGATGCGG 2249


RESULT   4
T03090
ID   T03090 standard; DNA; 4425 BP.
AC   T03090;
DT   14-FEB-1996  (first entry)
DE   Protein tyrosine-kinase SAL-S1 gene.
KW   Protein tyrosine-kinase; pTK; SAL-S1; agonist; cell growth;
KW   differentiation; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             30. .3927
FT                   /*tag=  a
FT   signal_peptide  30. .102
FT                   /*tag=  b
FT   mat_peptide     103. .3924
FT                   /*tag=  c
PN   WO9527061-A1.
PD   12-OCT-1995.
PF   04-APR-1995; U04228.
PR   04-APR-1994; US-222616.
PA   (GETH ) GENENTECH INC.
PI   Bennett BD,  Goeddel D,  Lee JM,  Matthews W,  Tsai SP;
PI   Wood WI;
DR   WPI; 95-366160/47.
DR   P-PSDB; R85937.
PT   Agonist antibodies which activate specific protein tyrosine
PT   kinase(s)  - also activate chimeric proteins of kinase extracellular
PT   domain and Ig constant domain, useful for studying, and therapeutic
PT   modulation of, cell growth and differentiation
PS   Disclosure; Fig 15A-F; 125pp; English.
CC   DNA probes based on protein tyrosine-kinase (pTK) sequences were used
CC   to screen cDNA libraries to identify novel pTK genes.  The SAL-S1
CC   gene (see also T03101) was isolated from several megakaryocytic cell
CC   libraries.  The gene can be used to produce recombinant SAL-S1 or its
CC   fragments, to detect related genes, and to design drugs, peptides
CC   or antisense nucleotides that modulate pTK activity.
SQ   Sequence   4425 BP;    939 A;    1348 C;    1361 G;    777 T;


  Query Match             0.9%;  Score 35.2;  DB 1;  Length 4425;
  Best Local Similarity   60.4%;  Pred. No. 2.7;
  Matches   58;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy   1875 acggcgactacaggcgggtcggggagcagtcagcagttttttcgagggatggcaagcgga 1934
          ||   ||| | |||| | | | |||  |||| | |||    | |  |||   ||| | ||
Db   2165 ACAAAGACGAGAGGCTGCTGGAGGAAAAGTCTGGAGTCGACTTGGCGGACTCCAACCAGA 2224

Qy   1935 cgggcagcgtgccacgagtgcgccaagcggaagcgg 1970
           |   ||| | |  || |||||| | | ||| ||||
Db   2225 AGCTGAGCATCCAGCGCGTGCGCGAGGAGGATGCGG 2260


RESULT   5
T03104
ID   T03104 standard; DNA; 9108 BP.
AC   T03104;
DT   14-FEB-1996  (first entry)
DE   Plasmid pRK5.tk1-1.1 encoding Sal S-1.
KW   Protein tyrosine-kinase; pTK; SAL-S1; agonist; cell growth;
KW   differentiation; pRK5.tk1-1.1; ss.
OS   Chimeric synthetic;
OS   Chimeric Homo sapiens.
PN   WO9527061-A1.
PD   12-OCT-1995.
PF   04-APR-1995; U04228.
PR   04-APR-1994; US-222616.
PA   (GETH ) GENENTECH INC.
PI   Bennett BD,  Goeddel D,  Lee JM,  Matthews W,  Tsai SP;
PI   Wood WI;
DR   WPI; 95-366160/47.
PT   Agonist antibodies which activate specific protein tyrosine
PT   kinase(s)  - also activate chimeric proteins of kinase extracellular
PT   domain and Ig constant domain, useful for studying, and therapeutic
PT   modulation of, cell growth and differentiation
PS   Example 7; Page 102-111; 125pp; English.
CC   PCR primers given in T03102-03 were used to amplify plasmid
CC   pRK5.tk1-1.1 (T03104) incorporating SAL-S1 (see T03090) DNA.
CC   The product was used to construct a fusion of the SAL-S1
CC   extracellular domain and human IgG1 Fc domain, which was expressed
CC   in 293 and COS7 cells and used to raise antibodies having
CC   protein tyrosine-kinase agonist activity.
SQ   Sequence   9108 BP;    2086 A;    2554 C;    2491 G;    1977 T;


  Query Match             0.9%;  Score 35.2;  DB 1;  Length 9108;
  Best Local Similarity   60.4%;  Pred. No. 4.6;
  Matches   58;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy   1875 acggcgactacaggcgggtcggggagcagtcagcagttttttcgagggatggcaagcgga 1934
          ||   ||| | |||| | | | |||  |||| | |||    | |  |||   ||| | ||
Db   3097 ACAAAGACGAGAGGCTGCTGGAGGAAAAGTCTGGAGTCGACTTGGCGGACTCCAACCAGA 3156

Qy   1935 cgggcagcgtgccacgagtgcgccaagcggaagcgg 1970
           |   ||| | |  || |||||| | | ||| ||||
Db   3157 AGCTGAGCATCCAGCGCGTGCGCGAGGAGGATGCGG 3192


RESULT   6
V21209_07/c
Continuation (8 of 17) of V21209 from base 700001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments  LOCUS V21209 Accession V21209
WP     Fragment Name      Begin      End
WP     V21209_00              1    110000
WP     V21209_01         100001    210000
WP     V21209_02         200001    310000
WP     V21209_03         300001    410000
WP     V21209_04         400001    510000
WP     V21209_05         500001    610000
WP     V21209_06         600001    710000
WP     V21209_07         700001    810000
WP     V21209_08         800001    910000
WP     V21209_09         900001   1010000
WP     V21209_10        1000001   1110000
WP     V21209_11        1100001   1210000
WP     V21209_12        1200001   1310000
WP     V21209_13        1300001   1410000
WP     V21209_14        1400001   1510000
WP     V21209_15        1500001   1610000
WP     V21209_16        1600001   1664976


  Query Match             0.8%;  Score 35;  DB 1;  Length 110000;
  Best Local Similarity   49.7%;  Pred. No. 32;
  Matches   89;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy   3823 ctaggaataagttgacttaaggctttaaaacgcctggaagtcattgacgcattcactatt 3882
          || |  | ||    |  || | |  || ||  ||    | | |||||| || |     |||
Db  10441 CTTGAGACAAAGATAGATAGGACAATATAAGCCCATATAATTATTGAAGCGTATGGAATT 10382

Qy   3883 ttatataaatatatacactattagggtccgcagcaacttacggttttaacacaagctgta 3942
          |  || | || |||| | || ||   |   |      | |  |||  |  ||||||| |
Db  10381 TCGTAGATATTTATAAAATAATATAATATACCAAGCATCATAGTTAAATAACAAGCTGAA 10322

Qy   3943 cgtatctcatctctagaattttgtgttagtttgtggacactaagtgtaacagctacgct 4001
           | ||   |||| |  ||| |    |||   | |      |  |   ||| ||  || |
Db  10321 GGAATAATATCTTTCCAATATCTCATTAAAATTTTATACATCGGAAAAACTGCGGCGAT 10263
```

```
RESULT   7
V86487/c
ID   V86487 standard; cDNA; 279 BP.
AC   V86487;
DT   27-APR-1999  (first entry)
DE   EST clone AT111.
KW   Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW   tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW   chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW   receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
OS   Homo sapiens.
PN   WO9845435-A2.
PD   15-OCT-1998.
PF   10-APR-1998; U06954.
PR   10-APR-1997; US-835913.
PA   (GEMY ) GENETICS INST INC.
PI   Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI   Racie LA, Spaulding V, Treacy M;
DR   WPI; 99-070076/06.
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries
PS   Claim 1; Page 251; 633pp; English.
CC   This sequence represents an expressed sequence tag (EST), and is a
CC   polynucleotide of the invention. The polynucleotides of the invention are
CC   all secreted EST sequences isolated from a variety of human tissue
CC   sources. The EST sequences and proteins encoded by them are predicted to
CC   have useful biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals, although no supporting data is given. Suggested activities
CC   include nutritional activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The EST sequences are also stated to be useful for gene
CC   therapy.
SQ   Sequence   279 BP;    77 A;    63 C;    36 G;    102 T;


  Query Match             0.8%;  Score 34.4;  DB 1;  Length 279;
  Best Local Similarity   54.4%;  Pred. No. 0.68;
  Matches   68;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy      43 aatagtttttggtatacggatagaatttggatgaaaaataaaacgaaatcaaaacatttt 102
           | ||||| |   || | ||  | | || ||  | |||| ||||||  || |||  | |||
Db     161 ATTAGTTCTGAATAAAAGGCAAAAGTTAGGTGGTAAAAAAAAACGGGATAAAATTAATTT 102

Qy     103 tcaaaagcgtggaagttttggccggcttgtgggcatggcaaaacgttttttggctatccg 162
           | ||  |  ||  ||| ||| |      ||  | ||||     | |  |||  |
Db     101 AGGAGAGAATAAAACCTTTTGCCTGTGAAAAGGTGTAGCAATTGAGTGTAAGGCAGTTTT 42

Qy     163 ttaat 167
           ||| |
Db      41 TTANT 37


RESULT   8
V54587
ID   V54587 standard; cDNA; 2447 BP.
AC   V54587;
DT   30-OCT-1998  (first entry)
DE   Human secretory protein encoding cDNA clone CO1020-1.
KW   Secretory protein; AM-973-1; BK260-2; BR390-1; CJ539-3; CN729-3;
KW   CO1390-3; CO1020-1; CS752-3; DM340-1; DW902-1; nutritional source;
KW   cytokine; cell proliferation; immune stimulation; immune suppression;
KW   haematopoiesis regulation; receptor/ligand activity; tumour inhibition;
KW   anti-inflammatory activity; cadherin; tumour invasion suppression;
KW   activin; inhibin; chemostatic; chemokinetic; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             184. .1191
FT                   /*tag= a
FT   CDS             347. .592
FT                   /*tag= b
FT                   /note= "encoded protein shown in W75861"
PN   WO9833916-A2.
PD   06-AUG-1998.
PF   30-JAN-1998; U01811.
PR   28-JAN-1998; US-014969.
PR   31-JAN-1997; US-792511.
PA   (GEMY ) GENETICS INST INC.
PI   Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI   Racie LA, Spaulding V, Treacy M;
DR   WPI; 98-437469/37.
DR   P-PSDB; W75857, W75861.
PT   New nucleic acid encoding secreted protein from human cells - and
PT   related vectors, transformed cells and proteins
PS   Claim 30; Pages 80-82; 114pp; English.
CC   Sequences shown in V54581 to V54590 represent cDNA sequences encoding
CC   human secretory proteins (W75851 to W75860). The sequences correspond to
CC   the clones designated AM-973-1, BK260-2, BR390-1, CJ539-3, CN729-3,
CC   CO1390-3, CO1020-1, CS752-3, DM340-1 and DW902-1 respectively. The
CC   polynucleotides and polypeptides can be used as e.g. nutritional sources
CC   or supplements. They may exhibit e.g. cytokine and cell proliferation/
CC   differentiation activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, receptor/ligand activity, anti-
CC   inflammatory activity, cadherin/tumour invasion suppressor activity,
CC   tumour inhibition activity. They may also exhibit other activities such
CC   as tissue growth activity, activin/inhibin activity and chemostatic/
CC   chemokinetic activity.
SQ   Sequence   2447 BP;    832 A;    458 C;    506 G;    650 T;


  Query Match             0.8%;  Score 34.4;  DB 1;  Length 2447;
  Best Local Similarity   48.4%;  Pred. No. 3.2;
  Matches   92;  Conservative    1;  Mismatches   97;  Indels    0;  Gaps    0;

Qy    3939 tgtacgtatctcatctctagaattttgtgttagtttgtggacactaagtgtaacagctac 3998
           | || |  || ||| ||| ||  | ||| || | || || ||  ||| ||| |||  ||
Db    2173 TATATGGTTCACATATCTGGATCTGTGTATTTGATTTTGTACTTTAAATGTGACAAATAA 2232

Qy    3999 gctccggtaggttaaggaactaaactaaatgaatcagatatatacacatatattttcgcg 4058
           |       |   ||  ||  |||  :||  ||  | | | | | | | | |
Db    2233 ACCTTTTGGGAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2292

Qy    4059 taattatataaactacatagtgtcttaaagcgcctcagcctaatataaaatgactaaatg 4118
           ||  | | |||  | | |       |||       |    || | ||||  |  |||
Db    2293 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2352

Qy    4119 ttaaaataaa 4128
           |||| |||
Db    2353 AAAAAAAAAA 2362


RESULT   9
V74330/c
ID   V74330 standard; DNA; 7363 BP.
AC   V74330;
DT   16-MAR-1999  (first entry)
DE   Staphylococcus aureus contig SEQ ID #19.
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
OS   Staphylococcus aureus.
FH   Key             Location/Qualifiers
FT   misc_feature    901. .960
FT                   /*tag= a
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    2701. .2760
```

```
FT                   /*tag= b
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    4501. .4560
FT                   /*tag= c
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    6301. .6360
FT                   /*tag= d
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
PN   EP-786519-A2.
PD   30-JUL-1997.
PF   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009861.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI; 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 267-271; 3271pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence   7363 BP;    2558 A;    1013 C;    1286 G;    2262 T;


  Query Match              0.8%;  Score 34.4;  DB 1;  Length 7363;
  Best Local Similarity   60.9%;  Pred. No. 7.1;
  Matches   56;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy   1986 aaaagctcctctagcgatcgatataaacgaaacggccgaggaaatgatgagacttcgata 2045
         |||||| | |  | ||  | || |  || |   || | | | ||| ||| || | |   |
Db   1690 AAAAGCCCTTGAATCGTCGCCATCTTCACCATTAGGACCATGTAATAATGGGAATACTGCA 1631

Qy   2046 tcccagcttcgacctaaactatttccctttcaa 2077
          ||  |   | | ||| ||||| || |||||||
Db   1630 TCGTATGGTTGTCCTGAACTACTTTCTTTCAA 1599


RESULT  10
V57903/c
ID   V57903 standard; DNA; 237326 BP.
AC   V57903;
DT   21-DEC-1998  (first entry)
DE   Hereditary haemochromatosis subregion from an HH affected individual.
KW   Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;
KW   diagnosis; iron metabolism; NPT3; NPT4; RoRet; BTF1; BTF2; BTF3;
KW   BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
KW   type 1 sodium transport gene; ss.
OS   Homo sapiens.
PN   WO9814466-A1.
PD   09-APR-1998.
PF   30-SEP-1997; U17658.
PR   07-MAY-1997; US-852495.
PR   01-OCT-1996; US-724394.
PA   (PROG-) PROGENTOR INC.
PI   Feder JN, Kronmal GS, Lauer PM, Ruddy DA, Thomas WJ,
PI   Tsuchihashi Z, Wolff RK;
DR   WPI; 98-240014/21.
PT   Hereditary haemochromatosis gene products - used to develop products
PT   for the diagnosis and treatment of hereditary disorders in iron
PT   metabolism
PS   Claim 1; Fig 9; 209pp; English.
CC   The present invention describes hereditary haemochromatosis gene
CC   products from the human haemochromatosis gene. The present sequence
CC   represents a hereditary haemochromatosis subregion from an hereditary
CC   haemochromatosis (HH) affected individual. Also described is a
CC   method to determine the presence or absence of the common hereditary
CC   haemochromatosis (HFE) gene mutation in an individual comprising:
CC   (a) providing DNA or RNA from the individual; and (b) assessing the
CC   DNA or RNA for the presence or absence of a haplotype or genotype where
CC   the presence or absence of the haplotype genotype indicates the likely
CC   presence of the HFE gene mutation in the genome of the individual. The
CC   HFE gene sequences from the present invention can be used to develop
CC   products for use in the diagnosis and treatment of HFE. The present
CC   invention also describes BTF genes, which are homologues of the milk
CC   protein butyrophilin (BT), and can be used in the production of agonists
CC   and antagonists of BT function. Also described are: (1) a RoRet gene
CC   which can be used to develop products for the study, diagnosis and
CC   treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
CC   which are homologues of a type 1 sodium transport gene, and can
CC   similarly be used for hypophosphatemia.
SQ   Sequence   237326 BP;    69596 A;    48904 C;    48217 G;    70609 T;


  Query Match              0.8%;  Score 34.4;  DB 1;  Length 237326;
  Best Local Similarity   47.8%;  Pred. No. 83;
  Matches  130;  Conservative    0;  Mismatches  141;  Indels    1;  Gaps    1;

Qy   3858 ggaagtcattgacgcattcactatttatataaatatatacactattagggtccgcagca 3917
         | ||   ||    | | || | |  ||| |    | |  |||  | ||  ||  ||   |
Db  58126 GCAAAGGATAATTGAAATCCCAAACTTACAAGGTTTTCAACAAAAGTAAAGTTTGCTAAA 58067

Qy   3918 acttacggttttaacacaagctgtacgtatctcatctctagaattttgtgttagtttgtg 3977
          | |||    | |||||     | ||   |  ||   ||| |     |  |  ||| |
Db  58066 AGTTAACAGTGTAACATGTATTATAGTAACTTCTAATCTTGTGGCCTTAGACAGTCTAGT 58007

Qy   3978 gacactaagtgtaacagctacgctccggtaggttaaggaactaaactaaatgaatcagat 4037
               |  |  || |  | ||||  || |    | ||   | | |      ||  | |
Db  58006 CCACAGACATAAAAGAAGTTCGCTTTGGAAAAGAATGGTTATCATCCTCGGAAAAAAAAA 57947

Qy   4038 atatacacatatattttcgcgtaattatataaacta-catagtgtcttaaagcgcctcag 4096
          |   | | | | |       | | |||||||||  |  ||  ||| |  ||   |||
Db  57946 AAGGAAATAAAGAGGAGGCAGAATTTATATAAAAAAGAATGTTGTATGGAAAATCCTTGT 57887

Qy   4097 cctaatataaaatgactaaatgttaaaataaa 4128
          ||| | ||||| | ||||  |||| ||| |||
Db  57886 CCTGAGATAAATTAACTAGTTGTTTAAAGAAA 57855


RESULT  11
Q39893/c
ID   Q39893 standard; DNA; 376 BP.
AC   Q39893;
DT   20-MAY-1993  (first entry)
DE   Expressed Sequence Tag human gene marker EST00207.
KW   expressed sequence tag; human genome project; chromosome;
KW   human gene sequencing; PCR mapping; somatic cell hybrids;
KW   sublocalisation; gene tagging; tissue typing.
OS   Synthetic.
PN   WO9300353-A.
PD   07-JAN-1993.
```

```
PF   19-JUN-1992; U05222.
PR   20-JUN-1991; US-716831.
PR   12-FEB-1992; US-837195.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Adams MD, Venter JC;
DR   WPI; 93-036325/04.
PT   Particular expressed sequence tags from human CDNA - corresponds
PT   to transcription prods. of genes, useful for tagging genes,
PT   mapping chromosomes and tissue typing
PS   Claim 3; Page 176; 199pp; English.
CC   This sequence represents an EST (expressed sequence tag) ESTs are markers
CC   for human genes actually transcribed in vivo. Unlike the random genomic
CC   DNA sequence tagged sites (STSs), ESTs point directly to expressed genes.
CC   The use of ESTs could facilitate the tagging of most expressed human
CC   genes within a few years at a fraction of the cost of complete genomic
CC   sequencing. Using PCR primers Q39419-Q39580 (sequences designed
CC   from the ESTs) sublocalisation of an EST can be achieved with panels of
CC   fragments from specific chromosomes or pools of large genomic clones in
CC   an analogous manner. This sequence represents EST00207.
SQ   Sequence   376 BP;    88 A;    99 C;    109 G;    79 T;


  Query Match             0.8%;  Score 34.2;  DB 1;  Length 376;
  Best Local Similarity   51.7%;  Pred. No. 0.97;
  Matches   78;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy   1083 cttatgtagcgtaatcaccacagccgggagcttggccatgtcccttggcttgtgtttctt 1142
          | | || | | | || |  || |||| || |  | | |||| |      |  | |
Db    167 CCTCTGCAACCTCAGCATCTGAGTCGGGTTCTGGCGCTTCTCCCAAAGGACTCTTCCCAT 108

Qy   1143 ctttggcctgacaatttcgctgcagtcaaaggacattttcccctaccttgtaatccttgt 1202
           |||| ||   | |||   ||||||   ||| ||    ||  || |  | | |||  |||
Db    107 TTTTGACCCCTCCATTCTTCTGCAGGTTAAGCACCAGATCGTCTTCGCTCTCATCACTGT 48

Qy   1203 gggattggaaaatagcttggtgatcacaaag 1233
           |  |||  ||  |  |  |  || |  ||
Db     47 CAGCCTGGGTAAGTGGATCATACTCCCCTAG 17


RESULT  12
Q59305/c
ID   Q59305 standard; cDNA; 376 BP.
AC   Q59305;
DT   16-MAR-1994  (first entry)
DE   Human brain Expressed Sequence Tag EST00207.
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
OS   Homo sapiens.
PN   WO9316178-A.
PD   19-AUG-1993.
PF   12-FEB-1993; U01294.
PR   12-FEB-1992; US-837195.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Adams MD, Moreno RF, Venter CJ;
DR   WPI; 93-272882/34.
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
PS   Example 1; Page 153-154; 500pp; English.
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST00207 has an "excellent" coding probability as evaluated using the
CC   coding-region prediction program CRM.
CC   See also Q59041-Q61440.
SQ   Sequence   376 BP;    88 A;    99 C;    109 G;    79 T;


  Query Match             0.8%;  Score 34.2;  DB 1;  Length 376;
  Best Local Similarity   51.7%;  Pred. No. 0.97;
  Matches   78;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy   1083 cttatgtagcgtaatcaccacagccgggagcttggccatgtcccttggcttgtgtttctt 1142
          | | || | | | || |  || |||| || |  | | |||| |      |  | |
Db    167 CCTCTGCAACCTCAGCATCTGAGTCGGGTTCTGGCGCTTCTCCCAAAGGACTCTTCCCAT 108

Qy   1143 ctttggcctgacaatttcgctgcagtcaaaggacattttcccctaccttgtaatccttgt 1202
           |||| ||   | |||   ||||||   ||| ||    ||  || |  | | |||  |||
Db    107 TTTTGACCCCTCCATTCTTCTGCAGGTTAAGCACCAGATCGTCTTCGCTCTCATCACTGT 48

Qy   1203 gggattggaaaatagcttggtgatcacaaag 1233
           |  |||  ||  |  |  |  || |  ||
Db     47 CAGCCTGGGTAAGTGGATCATACTCCCCTAG 17


RESULT  13
X07378/c
ID   X07378 standard; DNA; 3060 BP.
AC   X07378;
DT   07-JUN-1999  (first entry)
DE   Arabidopsis TTG1 gene mutant allele ttg1.1.
KW   TTG1 gene; transparent testa glabra 1 gene; signal transduction;
KW   WD40 repeat protein; transgenic plant; trichome; anthocyanin;
KW   pigment; mutant; ss.
OS   Arabidopsis thaliana.
FH   Key             Location/Qualifiers
FT   CDS             1041. .1991
FT                   /*tag=  a
FT   mutation        1989
FT                   /*tag=  b
FT                   /note= "C in wild-type allele, introduces a stop
FT                          codon"
PN   WO9900501-A1.
PD   07-JAN-1999.
PF   25-JUN-1998; G01861.
PR   25-JUN-1997; GB-013430.
PA   (PLAN-) PLANT BIOSCIENCE LTD.
PI   Gray JC, Walker AR;
DR   WPI; 99-095746/08.
PT   New isolated transparent testa glabra 1 gene - used for producing
PT   plants with modified traits, e.g. insect protection, chemical or
PT   fibre production, salt removal, pigmentation, taste or seedling
PT   growth
PS   Claim 12; Page -; 61pp; English.
CC   This is the nucleotide sequence of the Arabidopsis thaliana
CC   Landsberg erecta TTG1 gene (transparent testa, glabra 1 gene)
CC   mutant allele ttg1.1.  It contains a point mutation (C to T) at
CC   codon of 317 of the coding region, converting a glutamine codon to
CC   a stop codon.  The TTG1 gene (see also X07374) encodes a WD40
CC   repeat protein (see W97835) suggested to be a component of a signal
CC   transduction pathway that regulates expression or action of
CC   downstream transcription factors.  TTG1 is thought to act upstream
CC   of an Arabidopsis R homologue in the pathways leading to trichome
CC   differentiation and anthocyanin synthesis.  The TTG1 gene, mutant
CC   alleles (see X07376-80 and X07383), vectors, host cells, transgenic
CC   plants, and polypeptide expression products are claimed, plus
CC   methods of manipulating the phenotypic characteristics associated
CC   with TTG1, e.g. the number of trichomes on the aerial parts of the
CC   plant, number of trichomes on the root hairs, mucilage of the seeds,
CC   dormancy of the seeds, anthocyanin pigmentation, condensation of
CC   the tannins, number of stomata on hypocotyls.  In particular,
CC   insect protection, chemical production, climate tolerance, salt
CC   removal, fibre production, ornamental value, water and nutrient
CC   absorption, initiation of seed germination, pigmentation, taste, or
CC   speed of seedling growth can be modified.  Plants with the ttg1.1
CC   allele have a dwarf phenotype.
CC   (NB. the nucleotide sequence for allele ttg1.1 was produced from
CC   information provided in Fig 3 of the specification).
SQ   Sequence   3060 BP;    864 A;    599 C;    565 G;    1032 T;


  Query Match             0.8%;  Score 34;  DB 1;  Length 3060;
```

```
  Best Local Similarity   50.6%;  Pred. No. 5.1;
  Matches   82;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy   3963 ttgtgttagtttgtggacactaagtgtaacagctacgctccggtaggttaaggaaactaaa 4022
          || | |  |||| ||     |  | | |    |  |  |      |   | ||  |||
Db   2605 TTCTATATGTTTCTGAGGGCATACAGCAGAGTCAGCAGTAAACAATTCTCAGATACTTCG 2546

Qy   4023 ctaaatgaatcagatatatacacatatattttcgcgtaattatataaactacatagtgtc 4082
           |||||  ||  | ||  | ||  || ||  |  | || |  ||| || ||  | | | ||||
Db   2545 TTAAATTAACAATATTTGTATACGTAGCTAGTGGCTTCGTTAAATTAAGAATAGATTGTC 2486

Qy   4083 ttaaagcgcctcagcctaatataaaatgactaaatgttaaaa 4124
          | |||       || || |   |  |||  ||||  ||| ||
Db   2485 TCAAAAAAAAATAGACTGAAGAATTATGGGTAAAAATTAGAA 2444


RESULT  14
X07374/c
ID   X07374 standard; DNA; 3060 BP.
AC   X07374;
DT   07-JUN-1999  (first entry)
DE   Arabidopsis TTG1 gene encoding plant regulatory protein.
KW   TTG1 gene; transparent testa glabra 1 gene; signal transduction;
KW   WD40 repeat protein; transgenic plant; trichome; anthocyanin;
KW   pigment; ss.
OS   Arabidopsis thaliana.
FH   Key             Location/Qualifiers
FT   CDS             1041. .2075
FT                   /*tag=  a
PN   WO9900501-A1.
PD   07-JAN-1999.
PF   25-JUN-1998; G01861.
PR   25-JUN-1997; GB-013430.
PA   (PLAN-) PLANT BIOSCIENCE LTD.
PI   Gray JC, Walker AR;
DR   WPI; 99-095746/08.
DR   P-PSDB; W97835.
PT   New isolated transparent testa glabra 1 gene - used for producing
PT   plants with modified traits, e.g. insect protection, chemical or
PT   fibre production, salt removal, pigmentation, taste or seedling
PT   growth
PS   Claim 3; Fig 3; 61pp; English.
CC   This is the nucleotide sequence of the novel Arabidopsis thaliana
CC   Landsberg erecta TTG1 gene (transparent testa, glabra 1 gene),
CC   which encodes a WD40 repeat protein (see W97835) suggested to be a
CC   component of a signal transduction pathway which regulates
CC   expression or action of downstream transcription factors.  TTG1
CC   is thought to act upstream of an Arabidopsis R homologue in the
CC   pathways leading to trichome differentiation and anthocyanin
CC   synthesis.  The TTG1 locus was identified by complementation of a
CC   ttg1 mutant with genomic DNA and by sequencing the gene in several
CC   ttg1 mutant alleles.  The 1.6 kb transcript is present in all major
CC   organs.  Mutant alleles (see X07376-80 and X07383) of TTG1 are also
CC   claimed, as well as vectors, host cells, transgenic plants,
CC   polypeptide expression products and other related materials, plus
CC   methods of manipulating the phenotypic characteristics associated
CC   with TTG1, e.g. the number of trichomes on the aerial parts of the
CC   plant, number of trichomes on the root hairs, mucilage of the seeds,
CC   dormancy of the seeds, anthocyanin pigmentation, condensation of
CC   the tannins, number of stomata on hypocotyls.  In particular,
CC   insect protection, chemical production, climate tolerance, salt
CC   removal, fibre production, ornamental value, water and nutrient
CC   absorption, initiation of seed germination, pigmentation, taste, or
CC   speed of seedling growth can be modified.
SQ   Sequence   3060 BP;    864 A;    600 C;    565 G;    1031 T;


  Query Match             0.8%;  Score 34;  DB 1;  Length 3060;
  Best Local Similarity   50.6%;  Pred. No. 5.1;
  Matches   82;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy   3963 ttgtgttagtttgtggacactaagtgtaacagctacgctccggtaggttaaggaaactaaa 4022
          || | |  |||| ||     |  | | |    |  |  |      |   | ||  |||

Db   2605 TTCTATATGTTTCTGAGGGCATACAGCAGAGTCAGCAGTAAACAATTCTCAGATACTTCG 2546

Qy   4023 ctaaatgaatcagatatatacacatatattttcgcgtaattatataaactacatagtgtc 4082
           |||||  ||  | ||  | ||  || ||  |  | || |  ||| || ||  | | | ||||
Db   2545 TTAAATTAACAATATTTGTATACGTAGCTAGTGGCTTCGTTAAATTAAGAATAGATTGTC 2486

Qy   4083 ttaaagcgcctcagcctaatataaaatgactaaatgttaaaa 4124
          | |||       || || |   |  |||  ||||  ||| ||
Db   2485 TCAAAAAAAAATAGACTGAAGAATTATGGGTAAAAATTAGAA 2444


RESULT  15
X07376/c
ID   X07376 standard; DNA; 3060 BP.
AC   X07376;
DT   07-JUN-1999  (first entry)
DE   Arabidopsis TTG1 gene mutant allele ttg1.10.
KW   TTG1 gene; transparent testa glabra 1 gene; signal transduction;
KW   WD40 repeat protein; transgenic plant; trichome; anthocyanin;
KW   pigment; mutant; ss.
OS   Arabidopsis thaliana.
FH   Key             Location/Qualifiers
FT   CDS             1041. .2075
FT                   /*tag=  a
FT   mutation        997
FT                   /*tag=  b
FT                   /note= "G in wild-type allele"
PN   WO9900501-A1.
PD   07-JAN-1999.
PF   25-JUN-1998; G01861.
PR   25-JUN-1997; GB-013430.
PA   (PLAN-) PLANT BIOSCIENCE LTD.
PI   Gray JC, Walker AR;
DR   WPI; 99-095746/08.
DR   P-PSDB; W97835.
PT   New isolated transparent testa glabra 1 gene - used for producing
PT   plants with modified traits, e.g. insect protection, chemical or
PT   fibre production, salt removal, pigmentation, taste or seedling
PT   growth
PS   Claim 12; Page -; 61pp; English.
CC   This is the nucleotide sequence of the Arabidopsis thaliana
CC   Landsberg erecta TTG1 gene (transparent testa, glabra 1 gene)
CC   mutant allele ttg1.10.  It contains a point mutation (G to A) in
CC   the 5' untranslated region when compared to the wild-type TTG1
CC   gene (see X07374).  The gene encodes a WD40 repeat protein (see
CC   W97835) suggested to be a component of a signal transduction
CC   pathway that regulates expression or action of downstream
CC   transcription factors.  TTG1 is thought to act upstream of an
CC   Arabidopsis R homologue in the pathways leading to trichome
CC   differentiation and anthocyanin synthesis.  The TTG1 gene, mutant
CC   alleles (see X07376-80 and X07383), vectors, host cells, transgenic
CC   plants, and polypeptide expression products are claimed, plus
CC   methods of manipulating the phenotypic characteristics associated
CC   with TTG1, e.g. the number of trichomes on the aerial parts of the
CC   plant, number of trichomes on the root hairs, mucilage of the seeds,
CC   dormancy of the seeds, anthocyanin pigmentation, condensation of
CC   the tannins, number of stomata on hypocotyls.  In particular,
CC   insect protection, chemical production, climate tolerance, salt
CC   removal, fibre production, ornamental value, water and nutrient
CC   absorption, initiation of seed germination, pigmentation, taste, or
CC   speed of seedling growth can be modified.  The mutation found in
CC   tgg1.10 may affect the translation of RNA.  Plants having the
CC   ttg1.10 allele have more trichomes, different seed mucilage and
CC   less anthocyanin.
CC   (NB. the nucleotide sequence for allele ttg1.10 was produced from
CC   information provided in Fig 3 of the specification).
SQ   Sequence   3060 BP;    865 A;    600 C;    564 G;    1031 T;


  Query Match             0.8%;  Score 34;  DB 1;  Length 3060;
  Best Local Similarity   50.6%;  Pred. No. 5.1;
  Matches   82;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;
```

```
Qy   3963 ttgtgttagtttgtggacactaagtgtaacagctacgctccggtaggttaaggaactaaa 4022
          || | |  |||| ||    |  |  | |    |  |  |     |   | ||  |||
Db   2605 TTCTATATGTTTCTGAGGGCATACAGCAGAGTCAGCAGTAAACAATTCTCAGATACTTCG 2546

Qy   4023 ctaaatgaatcagatatatacacatatattttcgcgtaattatataaactacatagtgtc 4082
           ||||| ||  | || | || || ||  |  | || |  ||| || ||  | | | ||||
Db   2545 TTAAATTAACAATATTTGTATACGTAGCTAGTGGCTTCGTTAAATTAAGAATAGATTGTC 2486

Qy   4083 ttaaagcgcctcagcctaatataaaatgactaaatgttaaaa 4124
          | |||       || || |   |  |||  ||||  ||| ||
Db   2485 TCAAAAAAAAATAGACTGAAGAATTATGGGTAAAAATTAGAA 2444
```

Search completed: September 4, 2000, 00:58:07
Job time: 18062 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2000, 00:29:28 ; Search time 183.89 Seconds
(without alignments)
1474.373 Million cell updates/sec

Title: US-09-332-522B-3
Perfect score: 1971
Sequence: 1 cggcacgaggattaatgctg.....aaaaaaaaaaaaaaaaaaac 1971

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	46	2.3	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	38.6	2.0	2007	5	US-08-747-221B-36	Sequence 36, Appl
c	3	38.6	2.0	2007	5	US-08-747-221B-38	Sequence 38, Appl
	4	38.2	1.9	3238	6	PCT-US94-10080-5	Sequence 5, Appli
	5	38	1.9	936	4	US-08-557-309B-21	Sequence 21, Appl
	6	38	1.9	936	5	US-08-834-306-21	Sequence 21, Appl
c	7	37.6	1.9	7218	1	US-08-232-463-14	Sequence 14, Appl
	8	37	1.9	684	2	US-08-226-264-27	Sequence 27, Appl
	9	36.6	1.9	314	2	US-08-686-878A-25	Sequence 25, Appl
	10	36.4	1.8	1454	3	US-08-713-000-7	Sequence 7, Appli
	11	36.4	1.8	1454	4	US-08-975-316-7	Sequence 7, Appli
	12	36.4	1.8	1474	4	US-08-975-316-71	Sequence 71, Appl
	13	36.4	1.8	7859	1	US-07-854-596B-4	Sequence 4, Appli
	14	36.4	1.8	7859	3	US-08-450-905B-15	Sequence 15, Appl
	15	36.4	1.8	7859	5	US-07-982-759F-15	Sequence 15, Appl
	16	35.2	1.8	19124	4	US-08-487-826B-13	Sequence 13, Appl
	17	35	1.8	946	5	US-08-916-443A-15	Sequence 15, Appl
	18	34.8	1.8	972	1	US-07-915-934-1	Sequence 1, Appli
	19	34.8	1.8	972	1	US-08-325-743-1	Sequence 1, Appli
	20	34.8	1.8	1023	1	US-08-252-966B-16	Sequence 16, Appl
	21	34.8	1.8	1223	5	US-09-154-874-4	Sequence 4, Appli
	22	34.6	1.8	536	1	US-08-341-568-1	Sequence 1, Appli
	23	34.6	1.8	536	3	US-08-911-020-1	Sequence 1, Appli
	24	34.4	1.7	1307	4	US-08-960-022-17	Sequence 17, Appl
	25	34.4	1.7	1639	4	US-08-737-524B-1	Sequence 1, Appli
	26	34.4	1.7	1897	1	US-08-184-632-1	Sequence 1, Appli
	27	34.4	1.7	2143	3	US-08-656-177A-1	Sequence 1, Appli
	28	34	1.7	1776	5	US-08-655-352-10	Sequence 10, Appl
	29	34	1.7	2901	6	PCT-US96-00419-4	Sequence 4, Appli
	30	33.8	1.7	940	3	US-08-471-717-1	Sequence 1, Appli
	31	33.8	1.7	1602	2	US-08-530-950-3	Sequence 3, Appli
	32	33.8	1.7	2026	4	US-08-993-228-3	Sequence 3, Appli
	33	33.8	1.7	4843	5	US-08-986-485-1	Sequence 1, Appli
	34	33.6	1.7	1393	1	US-08-174-467-18	Sequence 18, Appl
	35	33.6	1.7	1393	5	US-08-452-071-18	Sequence 18, Appl
	36	33.6	1.7	1395	4	US-08-553-367A-1	Sequence 1, Appli
	37	33.6	1.7	1792	5	US-08-747-221B-18	Sequence 18, Appl
c	38	33.6	1.7	1792	5	US-08-747-221B-20	Sequence 20, Appl
	39	33.6	1.7	2550	7	5258287-23	Patent No. 5258287
	40	33.6	1.7	3527	4	US-08-909-965C-7	Sequence 7, Appli
	41	33.4	1.7	1687	1	US-08-143-219-26	Sequence 26, Appl
	42	33.4	1.7	1813	6	PCT-US94-12883-3	Sequence 3, Appli
	43	33.4	1.7	1898	1	US-08-342-411A-1	Sequence 1, Appli
	44	33.2	1.7	1137	2	US-08-706-214-2	Sequence 2, Appli
	45	33.2	1.7	1804	4	US-08-504-459-5	Sequence 5, Appli

ALIGNMENTS

```
RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14
```

```
  Query Match             2.3%;  Score 46;  DB 1;  Length 7218;
  Best Local Similarity   4.7%;  Pred. No. 0.0014;
  Matches   19;  Conservative  215;  Mismatches  170;  Indels    0;  Gaps    0;

Qy     153 aactgtagaatccgcatcaccatcctaatcccaacaaaccaatgacatcttgagacctca 212
           | ||   || ::: : :: :: :::: :::: :   ::  :  : ::::    ::::
Db    1057 AGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1116

Qy     213 ccagccatggatcccttcgtgttcttcatagtactggcatcgctttatggcgttctttac 272
           ::  :: :   ::::::: : :::::: :  : ::  : :: :::: :  : :::::: :
Db    1117 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1176

Qy     273 tttttcgaccgcttcttcaagagttgcatgcactacccgtacgatgccttcctcaagaac 332
           ::::::  :: :::::::      :: : : :: :: : :  : ::::::::     :
Db    1177 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1236

Qy     333 accgggctgagtataaatttcatgagcctccactggcacacgagtgcctttaacaggacc 392
            ::   ::   : :   :::: :   ::::: ::  : : :   : :::::  :    ::
Db    1237 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1296

Qy     393 ctcctacgctggggatctgccggtaacagctgcacccggagagtaatgatcaccagcttt 452
           ::::: : ::     ::: ::  :  :  :: : :::      :  :  :: ::  ::::
Db    1297 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1356

Qy     453 aatgtaggagtcctggtcacctttctctgctcccgatcggtctgatcctgctcattgcc 512
             : :     ::::  :: :::::::::: ::::::  ::  :::  :::: ::: :: ::
Db    1357 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1416

Qy     513 actatcttcagcagtggtgaacaagacagctcttcgtctgtatc 556
            :: :::::  :  :  :| || | |    |||   ||| || |
Db    1417 YYYYYYYYYYYYYYYYYYGTACCAAATTCTTCTATCTCTTTAAC 1460


RESULT   2
US-08-747-221B-36
; Sequence 36, Application US/08747221B
; Patent No. 6063610
;  GENERAL INFORMATION:
;    APPLICANT:  Silver, Gary W.
;    APPLICANT:  Wisnewski, Nancy
;    TITLE OF INVENTION:  No. 6063610el Carboxylesterase Nucleic Acid
;    TITLE OF INVENTION:  Molecules, Proteins and Uses Thereof
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carol Talkington Verser, Ph.D.
;      ADDRESSEE:  Heska Corporation
;      STREET:  1825 Sharp Point Drive
;      CITY:  Fort Collins
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80525
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  WordPerfect for Windows, Version 7.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/747,221B
;      FILING DATE:  No. 6063610ember 12, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verser, Carol Talkington
;      REGISTRATION NUMBER:  37,459
;      REFERENCE/DOCKET NUMBER:  FC-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  970/493-7272
;      TELEFAX:  970/484-9505
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2007 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  11..1594
US-08-747-221B-36


  Query Match             2.0%;  Score 38.6;  DB 5;  Length 2007;
  Best Local Similarity   61.4%;  Pred. No. 0.11;
  Matches   62;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy    1870 agctttttgtcatagtccttagtcttaactcgtatttatttttcgtacggttgtcgagctc 1929
           | |||| | ||||| |    |  | ||   || | ||| ||||  | |   | |   | |
Db    1907 ATCTTTGTATCATATTTGCTTTTATTTTTTCATTTTTTTTTTATTTCAAATATATTGTTT 1966

Qy    1930 aaaaataaaatcaaattaagctaaaaaaaaaaaaaaaaaaa 1970
               ||||||  |||  ||   ||||||||||||||||||
Db    1967 TTTTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2007


RESULT   3
US-08-747-221B-38/c
; Sequence 38, Application US/08747221B
; Patent No. 6063610
;  GENERAL INFORMATION:
;    APPLICANT:  Silver, Gary W.
;    APPLICANT:  Wisnewski, Nancy
;    TITLE OF INVENTION:  No. 6063610el Carboxylesterase Nucleic Acid
;    TITLE OF INVENTION:  Molecules, Proteins and Uses Thereof
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carol Talkington Verser, Ph.D.
;      ADDRESSEE:  Heska Corporation
;      STREET:  1825 Sharp Point Drive
;      CITY:  Fort Collins
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80525
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  WordPerfect for Windows, Version 7.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/747,221B
;      FILING DATE:  No. 6063610ember 12, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verser, Carol Talkington
;      REGISTRATION NUMBER:  37,459
;      REFERENCE/DOCKET NUMBER:  FC-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  970/493-7272
;      TELEFAX:  970/484-9505
;  INFORMATION FOR SEQ ID NO:  38:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2007 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-747-221B-38


  Query Match             2.0%;  Score 38.6;  DB 5;  Length 2007;
  Best Local Similarity   61.4%;  Pred. No. 0.11;
  Matches   62;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy    1870 agctttttgtcatagtccttagtcttaactcgtatttatttttcgtacggttgtcgagctc 1929
           | |||| | ||||| |    |  | ||   || | ||| ||||  | |   | |   | |
Db     101 ATCTTTGTATCATATTTGCTTTTATTTTTTCATTTTTTTTTTATTTCAAATATATTGTTT 42
```

```
Qy   1930 aaaaataaaatcaaattaagctaaaaaaaaaaaaaaaaaaa 1970
              |||||| ||| ||  ||||||||||||||||||
Db     41 TTTTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1

RESULT   4
PCT-US94-10080-5
; Sequence 5, Application PC/TUS9410080
;  GENERAL INFORMATION:
;    APPLICANT:  GENETICS INSTITUTE, INC.
;    TITLE OF INVENTION:  RECEPTOR PROTEINS
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute Inc.- Legal Affairs
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/10080
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/123,934
;      FILING DATE:  17-SEP-1993
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LAZAR, Steven R
;      REGISTRATION NUMBER:  32,618
;      REFERENCE/DOCKET NUMBER:  5203-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8260
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3238 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
;      CLONE:  CFK1-10a
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  474..2000
PCT-US94-10080-5


  Query Match             1.9%;  Score 38.2;  DB 6;  Length 3238;
  Best Local Similarity   59.8%;  Pred. No. 0.19;
  Matches   64;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy   1864 ttcgatagctttttgtcatagtccttagtcttaactcgtatttatttttcgtacggttgtc 1923
          || | |    | |  |  | ||  |||||  | | | ||||  |||     |   | |
Db   3094 TTTGTTTTAATCTACTTTTTGTATTTAGTAGTTATTTGTATAAATTAAATAAACTGTTTT 3153

Qy   1924 gagctcaaaaataaaatcaaattaagctaaaaaaaaaaaaaaaaaaa 1970
            |  ||||||| |||| |||  ||   ||||||||||||||||||
Db   3154 CAAGTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3200


RESULT   5
US-08-557-309B-21
; Sequence 21, Application US/08557309B
; Patent No. 5916572
;  GENERAL INFORMATION:
;    APPLICANT:  Reed,Steven G.
;    APPLICANT:  Skeiky, Yasir A.W.
;    APPLICANT:  Lodes, Michael J.
;    APPLICANT:  Houghton, Raymond L.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE DETECTION AND   PREVENTION OF
;    NUMBER OF SEQUENCES:  69
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/557,309B
;      FILING DATE:  14-NOV-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.422
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  936 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-557-309B-21


  Query Match             1.9%;  Score 38;  DB 4;  Length 936;
  Best Local Similarity   62.8%;  Pred. No. 0.1;
  Matches   59;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy   1877 tgtcatagtccttagtcttaactcgtatttatttttcgtacggttgtcgagctcaaaaata 1936
          ||| | ||  ||| ||  |   |  | | ||||||  |||||  |   ||  | | |||
Db    841 TGTGAAAGCTCTTTGTTCTTTTTTCTTTCTATTTTGAAACGGTGATTCCGCATATATATA 900

Qy   1937 aaatcaaattaagctaaaaaaaaaaaaaaaaaaa 1970
            |  |||  ||   ||||||||||||||||||
Db    901 TTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 934


RESULT   6
US-08-834-306-21
; Sequence 21, Application US/08834306
; Patent No. 6054135
;  GENERAL INFORMATION:
;    APPLICANT:  Reed,Steven G.
;    APPLICANT:  Skeiky, Yasir A.W.
;    APPLICANT:  Lodes, Michael J.
;    APPLICANT:  Houghton, Raymond L.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE DETECTION AND   PREVENTION OF
;    NUMBER OF SEQUENCES:  65
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/834,306
;      FILING DATE:  15-APR-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.422C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  936 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-834-306-21


  Query Match             1.9%;  Score 38;  DB 5;  Length 936;
  Best Local Similarity   62.8%;  Pred. No. 0.1;
  Matches   59;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy   1877 tgtcatagtccttagtcttaactcgtatttatttttcgtacggttgtcgagctcaaaaata 1936
          ||| | ||  ||| ||  |   |  | | ||||||   |||||  |   ||  | | |||
Db    841 TGTGAAAGCTCTTTGTTCTTTTTTCTTTCTATTTTGAAACGGTGATTCCGCATATATATA 900

Qy   1937 aaatcaaattaagctaaaaaaaaaaaaaaaaaaa 1970
            |  |||  ||   ||||||||||||||||||
Db    901 TTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 934


RESULT   7
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             1.9%;  Score 37.6;  DB 1;  Length 7218;
  Best Local Similarity   4.4%;  Pred. No. 0.47;
  Matches   16;  Conservative  191;  Mismatches  155;  Indels    0;  Gaps    0;

Qy    916 aacacgtagtggtcactgaactaacaaggaaatccccgctgaggggagagcgcggcttgc 975
          :: : : :: ::  :  :::  :: ::::::::     :  :::::::::: : ::   :
Db   1396 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1337

Qy    976 aagtggacaatcaaataacccaagtaaacggctgcccagtaaacagcgaggagagttggg 1035
          ::: ::: ::  ::: ::   ::: ::: ::  :   :: ::: :: ::::::::  :::
Db   1336 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1277

Qy   1036 tgacatgcctgcagaactctctgaagctcaagccgggctactgtgtgagtgcggacttcg 1095
           :: : :   : ::::      ::::  :::  :::  :  : : ::: : :::    :
Db   1276 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1217

Qy   1096 tgcagcttaacgacgaaagcagcgccatctcacatcatagcattgatggtcagctacagt 1155
           : ::   :: :: ::::: :: :  :    : :  : :: :  :: ::  ::  : ::
Db   1216 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1157

Qy   1156 gctgtgatgaactaaatccgaacgtaagctgcttcgaggtggtggaggacgcaaatggag 1215
          :  : :: :::  :::   ::: : :::  :    :::: :: :::::: : ::: ::::
Db   1156 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1097

Qy   1216 atgtgccggtggagctgccgcagcatgtatgtctcaatgtgcgccgcactttggaggagg 1275
          : : :  :: ::::  :  : :: : : :  || |||  | |  ||  ||   |   ||
Db   1096 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATCGCAAGCTCCCTCGACCTGCAGCCAAGC 1037

Qy   1276 tc 1277
          ||
Db   1036 TC 1035


RESULT   8
US-08-226-264-27
; Sequence 27, Application US/08226264
; Patent No. 5801017
;  GENERAL INFORMATION:
;    APPLICANT:  Werber, Moshe M.
;    APPLICANT:  Zeelon, Elisha P.
;    APPLICANT:  Levanon, Avigdor
;    APPLICANT:  Guy, Rachel
;    APPLICANT:  Goldlust, Arie
;    APPLICANT:  Rigbi, Meir
;    APPLICANT:  Panet, Amos
;    APPLICANT:  Fischer, Meir
;    TITLE OF INVENTION:  PRODUCTION OF RECOMBINANT FACTOR XA
;    TITLE OF INVENTION:  INHIBITORS OF LEECH HIRUDO MEDICINALIS
;    NUMBER OF SEQUENCES:  28
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/226,264
```

```
;       FILING DATE:  08-APR-94
;       CLASSIFICATION:  530
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  White, John P.
;       REGISTRATION NUMBER:  28,678
;       REFERENCE/DOCKET NUMBER:  40017-A/JPW/GJG
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212-278-0400
;       TELEFAX:  212-391-0525
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  27:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  684 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
US-08-226-264-27


  Query Match             1.9%;  Score 37;  DB 2;  Length 684;
  Best Local Similarity   71.0%;  Pred. No. 0.17;
  Matches   49;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy   1902 tatttattttcgtacggttgtcgagctcaaaaataaaatcaaattaagctaaaaaaaaaa 1961
          ||| | || |||| ||  | || |  |  |||| ||||  |||  ||  ||||||||||
Db    609 TATATTTTATCGTTCGCATATCAACATTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 668

Qy   1962 aaaaaaaaa 1970
          |||||||||
Db    669 AAAAAAAAA 677


RESULT   9
US-08-686-878A-25
; Sequence 25, Application US/08686878A
; Patent No. 5708157
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs, Kenneth
;    APPLICANT:  McCoy, John
;    APPLICANT:  LaVallie, Edward
;    APPLICANT:  Racie, Lisa
;    APPLICANT:  Merberg, David
;    APPLICANT:  Treacy, Maurice
;    APPLICANT:  Evans, Cheryl
;    APPLICANT:  Spaulding, Vikki
;    TITLE OF INVENTION:  SECRETED PROTEINS AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  ENCODING THEM
;    NUMBER OF SEQUENCES:  71
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/686,878A
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  314 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-686-878A-25


  Query Match             1.9%;  Score 36.6;  DB 2;  Length 314;
  Best Local Similarity   62.1%;  Pred. No. 0.14;
  Matches   54;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy   1884 gtccttagtcttaactcgtatttatttcgtacggttgtcgagctcaaaaataaaatcaa 1943
          | ||||| | | ||    ||    |||    | | || ||   || |||   |||| ||
Db    227 GGCCTTANTTTAAAAAATTAAANTTTTNANAAAGTTTTTCTTACTAAAAGGGAAAAAAAA 286

Qy   1944 attaagctaaaaaaaaaaaaaaaaaa 1970
          |  ||   ||||||||||||||||||
Db    287 AAAAAAAAAAAAAAAAAAAAAAAAAA 313


RESULT  10
US-08-713-000-7
; Sequence 7, Application US/08713000
; Patent No. 5850020
;  GENERAL INFORMATION:
;    APPLICANT:  Bloksberg, Leonard N.
;    APPLICANT:  Havukkala, Ilkka
;    APPLICANT:  Grierson, Alastair
;    TITLE OF INVENTION:  MATERIALS AND METHODS FOR THE
;    TITLE OF INVENTION:  MODIFICATION OF PLANT LIGNIN CONTENT
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Speckman Picard PLLC
;      STREET:  2601 Elliott Avenue, Suite 4185
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98121
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/713,000
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sleath, Janet
;      REGISTRATION NUMBER:  37,007
;      REFERENCE/DOCKET NUMBER:  11000.1003
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-269-0565
;      TELEFAX:  206-269-0563
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1454 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-713-000-7


  Query Match             1.8%;  Score 36.4;  DB 3;  Length 1454;
  Best Local Similarity   59.8%;  Pred. No. 0.41;
  Matches   61;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy   1869 tagctttttgtcatagtccttagtcttaactcgtatttatttcgtacggttgtcgagct 1928
          ||  |  |  ||  |   |||   |  || |  ||| | | ||||||   || |
Db   1351 TAAATAATAATCCAATGTCTTCTGCCAAATTAATATATGTATTCGTATTTTTATATGAAA 1410
```

```
Qy   1929 caaaaataaaatcaaattaagctaaaaaaaaaaaaaaaaaaa 1970
          ||||| ||||  |||  ||   ||||||||||||||||||||
Db   1411 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1452


RESULT  11
US-08-975-316-7
; Sequence 7, Application US/08975316
; Patent No. 5952486
;  GENERAL INFORMATION:
;    APPLICANT:  BLOKSBERG, Leonard N., HAVUKKALA, Ilkka
;    APPLICANT:  and GRIERSON, Alastair W.
;    TITLE OF INVENTION:  MATERIALS AND METHODS FOR
;    TITLE OF INVENTION:  THE MODIFICATION OF PLANT LIGNIN CONTENT
;    NUMBER OF SEQUENCES:  88
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Law Offices of Ann W. Speckman
;      STREET:  2601 Elliott Avenue, Suite 4185
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98121
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/975,316
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/713,000
;      FILING DATE:  September 11, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SLEATH, Janet
;      REGISTRATION NUMBER:  37,007
;      REFERENCE/DOCKET NUMBER:  11000/1003C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-269-0565
;      TELEFAX:  206-269-0563
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1454 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-975-316-7


  Query Match             1.8%;  Score 36.4;  DB 4;  Length 1454;
  Best Local Similarity   59.8%;  Pred. No. 0.41;
  Matches   61;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy   1869 tagctttttgtcatagtccttagtcttaactcgtatttatttttcgtacggttgtcgagct 1928
          ||  |  |  ||  |  |||  |  || | ||| | | ||||||  |||
Db   1351 TAAATAATAATCCAATGTCTTCTGCCAAATTAATATATGTATTCGTATTTTTATATGAAA 1410

Qy   1929 caaaaataaaatcaaattaagctaaaaaaaaaaaaaaaaaaa 1970
          ||||| ||||  |||  ||   ||||||||||||||||||||
Db   1411 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1452


RESULT  12
US-08-975-316-71
; Sequence 71, Application US/08975316
; Patent No. 5952486
;  GENERAL INFORMATION:
;    APPLICANT:  BLOKSBERG, Leonard N., HAVUKKALA, Ilkka
;    APPLICANT:  and GRIERSON, Alastair W.
;    TITLE OF INVENTION:  MATERIALS AND METHODS FOR
;    TITLE OF INVENTION:  THE MODIFICATION OF PLANT LIGNIN CONTENT
;    NUMBER OF SEQUENCES:  88
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Law Offices of Ann W. Speckman
;      STREET:  2601 Elliott Avenue, Suite 4185
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98121
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/975,316
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/713,000
;      FILING DATE:  September 11, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SLEATH, Janet
;      REGISTRATION NUMBER:  37,007
;      REFERENCE/DOCKET NUMBER:  11000/1003C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-269-0565
;      TELEFAX:  206-269-0563
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  71:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1474 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-975-316-71


  Query Match             1.8%;  Score 36.4;  DB 4;  Length 1474;
  Best Local Similarity   59.8%;  Pred. No. 0.41;
  Matches   61;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy   1869 tagctttttgtcatagtccttagtcttaactcgtatttatttttcgtacggttgtcgagct 1928
          ||  |  |  ||  |  |||  |  || | ||| | | ||||||  |||
Db   1371 TAAATAATAATCCAATGTCTTCTGCCAAATTAATATATGTATTCGTATTTTTATATGAAA 1430

Qy   1929 caaaaataaaatcaaattaagctaaaaaaaaaaaaaaaaaaa 1970
          ||||| ||||  |||  ||   ||||||||||||||||||||
Db   1431 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1472


RESULT  13
US-07-854-596B-4
; Sequence 4, Application US/07854596B
; Patent No. 5434073
;  GENERAL INFORMATION:
;    APPLICANT:  Dawson, Keith M
;    APPLICANT:  Hunter, Michael G
;    APPLICANT:  Czaplewski, Lloyd G
;    TITLE OF INVENTION:  Proteins and nucleic acids
;    NUMBER OF SEQUENCES:  73
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dr. John J. McDonnell
;      STREET:  Ten South Wacker Drive, Suite 3000
;      CITY:  Chicago
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/854,596B
;      FILING DATE:  03-JUN-1992
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McDonnell, John J
;      REGISTRATION NUMBER:  26,949
;      REFERENCE/DOCKET NUMBER:  92,337
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-715-1000
;      TELEFAX:  312-715-1234
;      TELEX:  910-221-5317
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7859 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..7859
;      OTHER INFORMATION:  /note= "sequence of plasmid pSW6"
US-07-854-596B-4


  Query Match             1.8%;  Score 36.4;  DB 1;  Length 7859;
  Best Local Similarity   66.7%;  Pred. No. 1.1;
  Matches   52;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy   1893 cttaactcgtatttatttttcgtacggttgtcgagctcaaaaataaaatcaaattaagcta 1952
          |||||   |  | | ||||  || | |  | |  |  |||||  ||||  |||  ||   |
Db   5575 CTTAAAAAGATTCTCTTTTTTTTATGATATTTGTACAAAAAAAAAAAAAAAAAAAAAAAAA 5634

Qy   1953 aaaaaaaaaaaaaaaaa 1970
          |||||||||||||||||
Db   5635 AAAAAAAAAAAAAAAAA 5652


RESULT  14
US-08-450-905B-15
; Sequence 15, Application US/08450905B
; Patent No. 5856301
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Stem Cell Inhibiting Proteins
;    NUMBER OF SEQUENCES:  178
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HALE and DORR
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  MA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/450,905B
;      FILING DATE:  26-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/982,759
;      FILING DATE:  08-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9127319.3
;      FILING DATE:  23-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9221587.0
;      FILING DATE:  14-OCT-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BAKER, HOLLIE L.
;      REGISTRATION NUMBER:  31,321
;      REFERENCE/DOCKET NUMBER:  102.378.120DV-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-526-6110
;      TELEFAX:  617-526-5000
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7859 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA
US-08-450-905B-15


  Query Match             1.8%;  Score 36.4;  DB 3;  Length 7859;
  Best Local Similarity   66.7%;  Pred. No. 1.1;
  Matches   52;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy   1893 cttaactcgtatttatttttcgtacggttgtcgagctcaaaaataaaatcaaattaagcta 1952
          |||||   |  | | ||||  || | |  | |  |  |||||  ||||  |||  ||   |
Db   5575 CTTAAAAAGATTCTCTTTTTTTTATGATATTTGTACAAAAAAAAAAAAAAAAAAAAAAAAA 5634

Qy   1953 aaaaaaaaaaaaaaaaa 1970
          |||||||||||||||||
Db   5635 AAAAAAAAAAAAAAAAA 5652


RESULT  15
US-07-982-759F-15
; Sequence 15, Application US/07982759F
; Patent No. 6057123
;  GENERAL INFORMATION:
;    APPLICANT:  CRAIG, Stewart
;    APPLICANT:  GEORGE, Michael
;    APPLICANT:  EDWARDS, Richard Mark
;    APPLICANT:  CZAPLEWSKI, Lloyd George
;    APPLICANT:  GILBERT, Richard
;    TITLE OF INVENTION:  Stem Cell Inhibiting Proteins
;    NUMBER OF SEQUENCES:  178
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HALE and DORR LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  MA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/982,759F
;      FILING DATE:  08-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9127319.3
;      FILING DATE:  23-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9221587.0
;      FILING DATE:  14-OCT-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BAKER, HOLLIE L.
;      REGISTRATION NUMBER:  31,321
;      REFERENCE/DOCKET NUMBER:  102378.120
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-526-6000
;      TELEFAX:  617-526-5000
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7859 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA
```

US-07-982-759F-15

Query Match 1.8%; Score 36.4; DB 5; Length 7859;
Best Local Similarity 66.7%; Pred. No. 1.1;
Matches 52; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

```
Qy   1893 cttaactcgtatttatttttcgtacggttgtcgagctcaaaaataaaatcaaattaagcta 1952
          |||||   |  | | ||||  || | |  | |  |  ||||| ||||  |||  ||   |
Db   5575 CTTAAAAAGATTCTCTTTTTTTATGATATTTGTACAAAAAAAAAAAAAAAAAAAAAAAAA 5634

Qy   1953 aaaaaaaaaaaaaaaaa 1970
          |||||||||||||||||
Db   5635 AAAAAAAAAAAAAAAAA 5652
```

Search completed: September 4, 2000, 00:30:21
Job time: 19991 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2000, 00:38:25 ; Search time 299.13 Seconds
(without alignments)
1648.543 Million cell updates/sec

Title: US-09-332-522B-3
Perfect score: 1971
Sequence: 1 cggcacgaggattaatgctg.....aaaaaaaaaaaaaaaaaaac 1971

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	41	2.1	1651	1	Q80218	Human NDF-alpha2b
2	41	2.1	2335	1	Q80216	Human proNDF-alpha
3	40.2	2.0	1586	1	Q47967	Rape acyl-ACP thio
4	39.6	2.0	1078	1	Q25785	Protein synthesis
5	38.6	2.0	2007	1	V40754	C. felis esterase,
c 6	38.6	2.0	2007	1	V40755	C. felis esterase,
7	38.2	1.9	3238	1	Q83532	Sequence encoding
8	38	1.9	936	1	T69171	Trypanosoma cruzi
9	37.8	1.9	1378	1	V59706	Human secreted pro
10	37.8	1.9	2323	1	V59524	Human secreted pro
11	37	1.9	684	1	Q72956	Clone pSP65-XaI-11
12	36.6	1.9	314	1	V09108	3' nucleotide sequ
13	36.6	1.9	1118	1	Q20263	Deg-1 gene. DNA se
14	36.6	1.9	2116	1	Q35987	Tomato hsp80 cDNA
15	36.6	1.9	3628	1	N60848	Plasmid sequence e
16	36.4	1.8	1454	1	V23913	Plant CAD enzyme D
17	36.4	1.8	1474	1	V23914	Plant CAD enzyme D
18	36.4	1.8	2353	1	V45981	A. thaliana sterol
19	36.4	1.8	7859	1	Q12154	Shuttle vector pSW
20	36.4	1.8	7859	1	Q44265	pSW6 for expressio
21	36.4	1.8	7984	1	Q25185	pSW6 expression ve
22	36	1.8	1285	1	Q86744	TIMP-3 metalloprot
23	36	1.8	2218	1	V84503	Human secreted pro
24	36	1.8	3828	1	V06566	Arabidopsis cellul
25	36	1.8	4358	1	V42653	Nucleotide sequenc
26	35.6	1.8	465	1	V34173	Human secreted pro
27	35.6	1.8	1099	1	V34238	Human secreted pro
28	35.6	1.8	1410	1	Q15022	Hyoscyamine 6 beta
29	35.6	1.8	10660	1	Q84793	Spinocerebellar at
30	35.4	1.8	106	1	V00430	3' fragment of clo
31	35.4	1.8	801	1	N91234	cDNA clone 8F3 of
c 32	35.2	1.8	1203	1	T64815	Tumour suppressor
33	35.2	1.8	19124	1	T72882	Plasmodium var-7 g
34	35	1.8	760	1	V84625	Human secreted pro
35	35	1.8	886	1	X51754	DNA encoding a hum
36	35	1.8	1322	1	X04402	Human secreted pro
37	35	1.8	1337	1	X04345	Human secreted pro
38	35	1.8	1583	1	V58758	Human secreted pro
39	35	1.8	2377	1	Q20360	Human pro-urokinas
40	35	1.8	2427	1	Q04107	Human pro-urokinas
41	35	1.8	3347	1	N80639	Sequence of cDNA e
42	34.8	1.8	863	1	T91302	Human H1075-1 secr
43	34.8	1.8	971	1	Q73760	RP-8 Programmed ce
44	34.8	1.8	971	1	T35704	Coding sequence fo
45	34.8	1.8	1023	1	T70132	Max-interacting pr

ALIGNMENTS

RESULT 1
Q80218
ID Q80218 standard; DNA; 1651 BP.
AC Q80218;
DT 31-AUG-1995 (first entry)
DE Human NDF-alpha2b clone 17 DNA.
KW Alpha; beta; neu differentiation factor; NDF; human; rat; p185-neu;
KW tyrosine phosphorylation; differentiation; phenotype; proliferation;
KW wound; tumour; epithelial tissue; breast; stomach;
KW gastrointestinal disease; Barrett's oesophagus;
KW (non-)cystic kidney disease; inflammatory bowel disease; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 20. .820
FT /*tag= a
FT /product= Human NDF-alpha2b
PN WO9428133-A.
PD 08-DEC-1994.
PF 23-MAY-1994; U05759.
PR 21-MAY-1993; US-066384.
PA (AMGE-) AMGEN INC.
PI Hu S, Koski RA, Liu N, Pierce GF, Sugarman BJ;
PI Wen D;
DR WPI; 95-022805/03.
DR P-PSDB; R68563.
PT New recombinant neu differentiation factors and corresp. DNA -
PT are used in the treatment of tumours, dermal wounds, and
PT gastrointestinal, kidney and inflammatory bowel diseases.
PS Claim 25; Page 139-141; 341pp; English.
CC The sequences given in Q80215-22 encode human neu differentiation
CC factors (NDF's). The peptides encoded by these cDNA clones isolated
CC from human tissues and cell lines, possess the ability to stimulate
CC human p185-neu tyrosine phosphorylation. These peptides have the
CC ability to induce a differentiated phenotype in certain cell lines and
CC can stimulate or inhibit proliferation of certain cell lines. The
CC NDF's can be used to treat wounds, tumours derived from epithelial
CC tissue of the breast, stomach etc., gastrointestinal disease, Barrett's
CC oesophagus, (non-)cystic kidney disease or inflammatory bowel disease.
CC These cDNA sequences may be used in a recombinant plasmid for the
CC recombinant production of the NDF's in a pro- or eukaryotic host cell.
SQ Sequence 1651 BP; 510 A; 460 C; 352 G; 329 T;

Query Match 2.1%; Score 41; DB 1; Length 1651;
Best Local Similarity 54.2%; Pred. No. 0.095;
Matches 83; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

```
Qy   1818 cttcacgaaacacacaaaaacacagcgaattgtagcacctcaaagattcgatagcttttt 1877
          ||  |  ||||  | | |||||||  || |              |||  |||   |
Db   1489 CTGTATAAAACCTAAATAAACACATAGATTCACCTGTAAAACTTTATTTTATATAATAAA 1548

Qy   1878 gtcatagtccttagtcttaactcgtatttattttcgtacggttgtcgagctcaaaaataa 1937
          ||  |   |||||   | |||   |  || | ||    | ||| |  |    |||||  ||
Db   1549 GTATTCCACCTTAAATTAAACAATTTATTTTATTTTAGCAGTTCTGCAAAAAAAAAAAAA 1608

Qy   1938 aatcaaattaagctaaaaaaaaaaaaaaaaaaa 1970
```

```
          ||  |||  ||   ||||||||||||||||||||
Db   1609 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1641


RESULT   2
Q80216
ID   Q80216 standard; DNA; 2335 BP.
AC   Q80216;
DT   31-AUG-1995  (first entry)
DE   Human proNDF-alpha2b DNA.
KW   Alpha; beta; neu differentiation factor; NDF; human; rat; p185-neu;
KW   tyrosine phosphorylation; differentiation; phenotype; proliferation;
KW   wound; tumour; epithelial tissue; breast; stomach;
KW   gastrointestinal disease; Barrett's oesophagus;
KW   (non-)cystic kidney disease; inflammatory bowel disease; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             117. .1505
FT                   /*tag= a
FT                   /product= Human proNDF-alpha2b
PN   WO9428133-A.
PD   08-DEC-1994.
PF   23-MAY-1994; U05769.
PR   21-MAY-1993; US-066384.
PA   (AMGE-) AMGEN INC.
PI   Hu S,  Koski RA,  Liu N,  Pierce GF,  Sugarman BJ;
PI   Wen D;
DR   WPI; 95-022805/03.
DR   P-PSDB; R68561.
PT   New recombinant neu differentiation factors and corresp. DNA -
PT   are used in the treatment of tumours, dermal wounds, and
PT   gastrointestinal, kidney and inflammatory bowel diseases.
PS   Claim 25; Page 129-132; 341pp; English.
CC   The sequences given in Q80215-22 encode human neu differentiation
CC   factors (NDF's).  The peptides encoded by these cDNA clones isolated
CC   from human tissues and cell lines, possess the ability to stimulate
CC   human p185-neu tyrosine phosphorylation.  These peptides have the
CC   ability to induce a differentiated phenotype in certain cell lines and
CC   can stimulate or inhibit proliferation of certain cell lines.  The
CC   NDF's can be used to treat wounds, tumours derived from epithelial
CC   tissue of the breast, stomach etc., gastrointestinal disease, Barrett's
CC   oesophagus, (non-)cystic kidney disease or inflammatory bowel disease.
CC   These cDNA sequences may be used in a recombinant plasmid for the
CC   recombinant production of the NDF's in a pro- or eukaryotic host cell.
SQ   Sequence   2335 BP;    725 A;    627 C;    523 G;    460 T;


  Query Match             2.1%;  Score 41;  DB 1;  Length 2335;
  Best Local Similarity   54.2%;  Pred. No. 0.11;
  Matches   83;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy   1818 cttcacgaaacacacaaaaacacagcgaattgtagcacctcaaagattcgatagcttttt 1877
          ||  |  ||||  | | |||||||  || |               |||  |||   |
Db   2174 CTGTATAAAACCTAAATAAACACATAGATTCACCTGTAAAACTTTATTTTATATAATAAA 2233

Qy   1878 gtcatagtccttagtcttaactcgtatttatttttcgtacggttgtcgagctcaaaaataa 1937
          ||  |   |||||    | |||   |  || | ||    | ||| |  |    ||||| ||
Db   2234 GTATTCCACCTTAAATTAAACAATTTATTTTATTTTAGCAGTTCTGCAAAAAAAAAAAAA 2293

Qy   1938 aatcaaattaagctaaaaaaaaaaaaaaaaaaa 1970
          ||  |||  ||   ||||||||||||||||||||
Db   2294 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2326


RESULT   3
Q47967
ID   Q47967 standard; cDNA; 1586 BP.
AC   Q47967;
DT   25-MAR-1994  (first entry)
DE   Rape acyl-ACP thioesterase (Clone pNL3) coding sequence.
KW   Enzyme precursor; acyl-ACP thioesterase; plant; transformation;
KW   fatty acid; yield; composition; seed oil; Brassica napus; ss.
OS   Brassica napus.
FH   Key             Location/Qualifiers
FT   cds             169. .1269
FT                   /*tag= a
FT                   /product= Enzyme precursor.
FT                   /note= "Has acyl-ACP thioester activity."
PN   WO9318158-A.
PD   16-SEP-1993.
PF   03-MAR-1993; G00432.
PR   03-MAR-1992; GB-004583.
PA   (UNIL ) UNILEVER NV.
PA   (UNIL ) UNILEVER PLC.
PI   Hellyer SA,  Loader NM,  Safford R,  Slabas AR;
DR   WPI; 93-303472/38.
DR   P-PSDB; R41674.
PT   cDNA encoding enzyme with acyl-ACP-thio:esterase activity -
PT   useful for transforming plants, for altering properties e.g. seed
PT   storage oil composition
PS   Claim 2; Figure 2; 48pp; English.
CC   The nucleotide sequence is useful for transforming plant cells and
CC   altering that plants characteristics.  Characteristics which may be
CC   altered are (1) altering the level of acyl-ACP thioesterase
CC   activity; (2) fatty acid yield and/or fatty acid composition and (3)
CC   yield and/or composition of seed storage oil.
SQ   Sequence   1586 BP;    491 A;    324 C;    339 G;    432 T;


  Query Match             2.0%;  Score 40.2;  DB 1;  Length 1586;
  Best Local Similarity   60.6%;  Pred. No. 0.16;
  Matches   66;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy   1862 gattcgatagctttttgtcatagtccttagtcttaactcgtatttatttttcgtacggttg 1921
          ||| || | |  | || |||||  |   | | ||   ||  |||   | |  |  | ||
Db   1460 GATCCGCTTGTATCTTTTCATAAACTAGATTTTTTCATCCAATTATATATGCTGAGTTTC 1519

Qy   1922 tcgagctcaaaaataaaatcaaattaagctaaaaaaaaaaaaaaaaaaa 1970
          ||      ||||| ||||  |||  ||   ||||||||||||||||||
Db   1520 TCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1568


RESULT   4
Q25785
ID   Q25785 standard; DNA; 1078 BP.
AC   Q25785;
DT   10-DEC-1992  (first entry)
DE   Protein synthesis inhibiting gene.
KW   Plant pathogens; pathogen resistance; medicine; ss.
OS   Hordeum vulgare.
FH   Key             Location/Qualifiers
FT   cds             43. .888
FT                   /*tag= a
FT   5'utr           1. .42
FT                   /*tag= b
FT   3'utr           886. .1078
FT                   /*tag= c
FT   polya_signal    930. .935
FT                   /*tag= d
FT                   /note= "possible polyA signal"
FT   polya_signal    963. .976
FT                   /*tag= e
FT                   /note= "possible polyA signal"
FT   polya_signal    1002. .1011
FT                   /*tag= f
FT                   /note= "possible polyA signal"
FT   polya_site      1032. .1078
FT                   /*tag= g
PN   DE4040954-A.
PD   25-JUN-1992.
PF   20-DEC-1990; 040954.
PR   20-DEC-1990; DE-040954.
PA   (PLAC ) MAX PLANCK GES FORDERUNG WISSE.
PI   Jach G, Logemann J, Mundy J, Schell J;
DR   WPI; 92-217825/27.
DR   P-PSDB; R25048.
```

```
PT   Pathogen resistant plants prodn. using protein synthesis inhibiting
PT   gene - or its fusion products, inserted into the genome, also
PT   pharmaceutical use of derived synthesis inhibiting protein
PS   Disclosure; Fig 3; 23pp; German.
CC   The sequence is that of a gene encoding a protein which effectively
CC   blocks protein synthesis by plant pathogens, e.g. Trichoderma reesii
CC   and Fusarium sporotrichoides. It can be fused to active promoters
CC   such as wun-1 and the fusion introduced into plant genetic material
CC   to impart pathogen resistance to both mono- and di-cotyledonous
CC   plants. The gene was isolated from mature barley seeds.
SQ   Sequence   1078 BP;    276 A;    310 C;    321 G;    171 T;


  Query Match             2.0%;  Score 39.6;  DB 1;  Length 1078;
  Best Local Similarity   63.8%;  Pred. No. 0.19;
  Matches   60;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy   1877 tgtcatagtccttagtcttaactcgtatttatttttcgtacggttgtcgagctcaaaaata 1936
            ||||| ||   |   | | |||   ||  ||  |   ||    |||| || | ||||| |
Db    979 TGTCACAGAGTTACATATAAACAAATAAATAAATAATTAAAAATGTCCAGTTTAAAAAAA 1038

Qy   1937 aaatcaaattaagctaaaaaaaaaaaaaaaaaa 1970
           |||  |||  ||   ||||||||||||||||||
Db   1039 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1072


RESULT   5
V40754
ID   V40754 standard; cDNA; 2007 BP.
AC   V40754;
DT   23-SEP-1998  (first entry)
DE   C. felis esterase, nFE92007, coding sequence.
KW   Esterase; flea; protective immune response; carboxylesterase; arthropod;
KW   haematophagous ectoparasite infestation; nfE92007; ds.
OS   Ctenocephalides felis.
FH   Key             Location/Qualifiers
FT   CDS             11. .1597
FT                   /*tag=  a
PN   WO9821324-A1.
PD   22-MAY-1998.
PF   10-NOV-1997; U20598.
PR   12-NOV-1996; US-747221.
PA   (HESK-) HESKA CORP.
PI   Brandt KS, Silver GM, Wisnewski N;
DR   WPI; 98-297929/26.
DR   P-PSDB; W57862.
PT   New nucleic acid encoding carboxyl:esterase(s) from fleas - useful,
PT   e.g. in vaccines, for preventing infestation by haematophagous
PT   ectoparasites, particularly on cats and dogs
PS   Claim 1; Page 160-162; 230pp; English.
CC   This sequence encodes the flea esterase protein, nfE92007 (the
CC   complementary strand is shown in V40755), of the invention. When
CC   administered to animals, the protein induces a protective immune
CC   (antibody) response against carboxylesterase (CE), so they, or
CC   compositions containing CE proteins, are used, therapeutically or as
CC   vaccines, to protect particularly mammals and birds, specifically cats
CC   and dogs, against haematophagous ectoparasite infestation (HEP),
CC   specifically fleas. More generally the compositions can be used to treat
CC   arthropods generally, including pests of agricultural crops, trees,
CC   stored goods etc., also those that are vectors of disease. Fragments of
CC   the DNA can be used as probes and primers for identification or
CC   production of nucleic acid. Antibodies against the protein can be used
CC   for passive immunisation; to screen expression libraries; to isolate the
CC   protein and to target cytotoxic compounds to HEP. The compounds
CC   containing CE are effective against both adult and larval stages; they
CC   target CEs, including juvenile hormone, that are involved in development,
CC   metamorphosis, feeding, digestion and reproduction.
SQ   Sequence   2007 BP;    695 A;    286 C;    371 G;    655 T;


  Query Match             2.0%;  Score 38.6;  DB 1;  Length 2007;
  Best Local Similarity   61.4%;  Pred. No. 0.49;
  Matches   62;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy   1870 agctttttgtcatagtccttagtcttaactcgtatttatttttcgtacggttgtcgagctc 1929
            | |||| | ||||| |   |  | | ||   || | ||| ||||  | |   | |   | |
Db   1907 ATCTTTGTATCATATTTGCTTTTATTTTTTCATTTTTTTTTTTATTTCAAATATATTGTTT 1966

Qy   1930 aaaaataaaatcaaattaagctaaaaaaaaaaaaaaaaaa 1970
                |||||| |||  ||   ||||||||||||||||||
Db   1967 TTTTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2007


RESULT   6
V40755/c
ID   V40755 standard; cDNA; 2007 BP.
AC   V40755;
DT   23-SEP-1998  (first entry)
DE   C. felis esterase, nFE92007, coding sequence complementary strand.
KW   Esterase; flea; protective immune response; carboxylesterase; arthropod;
KW   haematophagous ectoparasite infestation; nfE92007; ds.
OS   Ctenocephalides felis.
PN   WO9821324-A1.
PD   22-MAY-1998.
PF   10-NOV-1997; U20598.
PR   12-NOV-1996; US-747221.
PA   (HESK-) HESKA CORP.
PI   Brandt KS, Silver GM, Wisnewski N;
DR   WPI; 98-297929/26.
PT   New nucleic acid encoding carboxyl:esterase(s) from fleas - useful,
PT   e.g. in vaccines, for preventing infestation by haematophagous
PT   ectoparasites, particularly on cats and dogs
PS   Claim 1; Page 165; 230pp; English.
CC   This sequence is the complementary strand of the DNA encoding the flea
CC   esterase protein, nfE92007 (see V40754 for coding strand), of the
CC   invention. When administered to animals, the protein induces a protective
CC   immune (antibody) response against carboxylesterase (CE), so they, or
CC   compositions containing CE proteins, are used, therapeutically or as
CC   vaccines, to protect particularly mammals and birds, specifically cats
CC   and dogs, against haematophagous ectoparasite infestation (HEP),
CC   specifically fleas. More generally the compositions can be used to treat
CC   arthropods generally, including pests of agricultural crops, trees,
CC   stored goods etc., also those that are vectors of disease. Fragments of
CC   the DNA can be used as probes and primers for identification or
CC   production of nucleic acid. Antibodies against the protein can be used
CC   for passive immunisation; to screen expression libraries; to isolate the
CC   protein and to target cytotoxic compounds to HEP. The compounds
CC   containing CE are effective against both adult and larval stages; they
CC   target CEs, including juvenile hormone, that are involved in development,
CC   metamorphosis, feeding, digestion and reproduction.
SQ   Sequence   2007 BP;    655 A;    371 C;    286 G;    695 T;


  Query Match             2.0%;  Score 38.6;  DB 1;  Length 2007;
  Best Local Similarity   61.4%;  Pred. No. 0.49;
  Matches   62;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy   1870 agctttttgtcatagtccttagtcttaactcgtatttatttttcgtacggttgtcgagctc 1929
            | |||| | ||||| |   |  | | ||   || | ||| ||||  | |   | |   | |
Db    101 ATCTTTGTATCATATTTGCTTTTATTTTTTCATTTTTTTTTTTATTTCAAATATATTGTTT 42

Qy   1930 aaaaataaaatcaaattaagctaaaaaaaaaaaaaaaaaa 1970
                |||||| |||  ||   ||||||||||||||||||
Db     41 TTTTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1


RESULT   7
Q83532
ID   Q83532 standard; DNA; 3238 BP.
AC   Q83532;
DT   28-SEP-1995  (first entry)
DE   Sequence encoding serine/threonine kinase receptor CFK1-10a.
KW   Bone morphogenic protein; receptor; serine/threonine kinase; BMP;
KW   bone; cartilage; injury; treatment; inhibition; ss.
OS   Rattus rattus.
FH   Key             Location/Qualifiers
```

```
FT   cds             474. .2003
FT                   /*tag= a
FT                   /product= Serine/threonine kinase receptor.
PN   WO9507982-A.
PD   23-MAR-1995.
PF   07-SEP-1994; U10080.
PR   17-SEP-1993; US-123934.
PA   (GEMY ) GENETICS INST INC.
PI   Celeste AJ,  Thies RS,  Wozney JM,  Yamaji N;
DR   WPI; 95-131350/17.
DR   P-PSDB; R70239.
PT   Truncated BMP and serine/threonine kinase receptor proteins -
PT   used to inhibit the effects of BMP-2 and/or BMP-4.
PS   Claim 3; Page 56-58; 83pp; English.
CC   Truncated bone morphogenic protein (BMP) receptors and
CC   serine/threonine kinase receptors may be used in compositions to
CC   inhibit the effects of BMP's such as BMP-2 and BMP-4.  The truncated
CC   receptors pref. comprise the ligand binding domain, but not the
CC   serine/threonine kinase and transmembrane domains.  The truncated
CC   proteins are soluble and will be excreted into supernatant by
CC   recombinant mammalian cells expressing them.  Such cells can be
CC   delivered in a medium or matrix which partially impedes their
CC   mobility, thereby localising the cells to a site of bone or
CC   cartilage injury.
SQ   Sequence   3238 BP;    852 A;    749 C;    776 G;    861 T;


  Query Match             1.9%;  Score 38.2;  DB 1;  Length 3238;
  Best Local Similarity   59.8%;  Pred. No. 0.79;
  Matches   64;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy    1864 ttcgatagctttttgtcatagtccttagtcttaactcgtatttatttttcgtacggttgtc 1923
           || | |    | |  |  | ||  |||||  | | | ||||  |||     |    | |
Db    3094 TTTGTTTTAATCTACTTTTTGTATTTAGTAGTTATTTGTATAAATTAAATAAACTGTTTT 3153

Qy    1924 gagctcaaaaataaaatcaaattaagctaaaaaaaaaaaaaaaaaaa 1970
             |  ||||||| ||||  |||  ||   ||||||||||||||||||
Db    3154 CAAGTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3200


RESULT   8
T69171
ID   T69171 standard; cDNA; 936 BP.
AC   T69171;
DT   12-JAN-1998  (first entry)
DE   Trypanosoma cruzi antigen cDNA.
KW   Antigen; epitope; vaccine; protective immunity; Chagas disease;
KW   diagnosis; therapy; immunoassay; ss.
OS   Trypanosoma cruzi Tulahean strain C2.
FH   Key             Location/Qualifiers
FT   CDS             2. .703
FT                   /*tag= a
PN   WO9718475-A1.
PD   22-MAY-1997.
PF   14-NOV-1996; U18624.
PR   14-NOV-1995; US-557309.
PA   (CORI-) CORIXA CORP.
PI   Houghton RL,  Lodes MJ,  Reed SG,  Skeiky YAW;
DR   WPI; 97-289413/26.
DR   P-PSDB; W26533.
PT   Diagnosing Trypanosoma cruzi infection by detecting antibodies to
PT   novel antigens - which are useful in vaccines to provide protective
PT   immunity against Chagas' disease
PS   Claim 1; Page 49-50; 110pp; English.
CC   This novel DNA molecule was isolated by screening a Trypanosoma
CC   cruzi cDNA expression library with pools of sera from infected
CC   individuals.  It encodes an antigenic polypeptide (see W26533).
CC   22 Isolated genomic DNA of cDNA molecules (T69151-72) encode full-
CC   length antigens (see W26530-41), or epitope-containing repeat
CC   sequences (see W19094-102, W19079-86 and W26542-44) of native
CC   antigens, that can be used in a variety of immunoassays for
CC   detecting T. cruzi infection in a blood, serum, plasma, saliva,
CC   cerebrospinal fluid or urine sample.  The  polypeptides are also
CC   useful in vaccines and pharmaceutical compositions for inducing
CC   protective immunity against Chagas disease.  The isolated DNA
CC   molecules can be used for recombinant production of the antigenic
CC   polypeptides.
SQ   Sequence   936 BP;    244 A;    255 C;    272 G;    165 T;


  Query Match             1.9%;  Score 38;  DB 1;  Length 936;
  Best Local Similarity   62.8%;  Pred. No. 0.49;
  Matches   59;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy    1877 tgtcatagtccttagtcttaactcgtatttatttttcgtacggttgtcgagctcaaaaata 1936
           ||| | ||  ||| ||  |   |  | | ||||||   |||||  |   ||  | | |||
Db     841 TGTGAAAGCTCTTTGTTCTTTTTTCTTTCTATTTTGAAACGGTGATTCCGCATATATATA 900

Qy    1937 aaatcaaattaagctaaaaaaaaaaaaaaaaaaa 1970
             |  |||  ||   ||||||||||||||||||
Db     901 TTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 934


RESULT   9
V59706
ID   V59706 standard; DNA; 1378 BP.
AC   V59706;
DT   19-JAN-1999  (first entry)
DE   Human secreted protein gene 14 clone HPMFD84.
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS   Homo sapiens.
PN   WO9839448-A2.
PD   11-SEP-1998.
PF   06-MAR-1998; U04493.
PR   02-OCT-1997; US-061060.
PR   07-MAR-1997; US-038621.
PR   07-MAR-1997; US-040161.
PR   07-MAR-1997; US-040162.
PR   07-MAR-1997; US-040163.
PR   07-MAR-1997; US-040333.
PR   07-MAR-1997; US-040334.
PR   07-MAR-1997; US-040336.
PR   07-MAR-1997; US-040626.
PR   11-APR-1997; US-043311.
PR   11-APR-1997; US-043312.
PR   11-APR-1997; US-043313.
PR   11-APR-1997; US-043314.
PR   11-APR-1997; US-043568.
PR   11-APR-1997; US-043569.
PR   11-APR-1997; US-043576.
PR   11-APR-1997; US-043578.
PR   11-APR-1997; US-043580.
PR   11-APR-1997; US-043669.
PR   11-APR-1997; US-043670.
PR   11-APR-1997; US-043671.
PR   11-APR-1997; US-043672.
PR   11-APR-1997; US-043674.
PR   23-MAY-1997; US-047492.
PR   23-MAY-1997; US-047500.
PR   23-MAY-1997; US-047501.
PR   23-MAY-1997; US-047502.
PR   23-MAY-1997; US-047503.
PR   23-MAY-1997; US-047581.
PR   23-MAY-1997; US-047582.
PR   23-MAY-1997; US-047583.
PR   23-MAY-1997; US-047584.
PR   23-MAY-1997; US-047585.
PR   23-MAY-1997; US-047586.
PR   23-MAY-1997; US-047587.
```

PR 23-MAY-1997; US-047588.
PR 23-MAY-1997; US-047589.
PR 23-MAY-1997; US-047590.
PR 23-MAY-1997; US-047592.
PR 23-MAY-1997; US-047593.
PR 23-MAY-1997; US-047594.
PR 23-MAY-1997; US-047595.
PR 23-MAY-1997; US-047596.
PR 23-MAY-1997; US-047597.
PR 23-MAY-1997; US-047598.
PR 23-MAY-1997; US-047599.
PR 23-MAY-1997; US-047600.
PR 23-MAY-1997; US-047601.
PR 23-MAY-1997; US-047612.
PR 23-MAY-1997; US-047613.
PR 23-MAY-1997; US-047614.
PR 23-MAY-1997; US-047615.
PR 23-MAY-1997; US-047617.
PR 23-MAY-1997; US-047618.
PR 23-MAY-1997; US-047632.
PR 23-MAY-1997; US-047633.
PR 06-JUN-1997; US-048964.
PR 06-JUN-1997; US-048974.
PR 13-JUN-1997; US-049610.
PR 08-JUL-1997; US-051926.
PR 16-JUL-1997; US-052874.
PR 18-AUG-1997; US-055724.
PR 22-AUG-1997; US-056630.
PR 22-AUG-1997; US-056631.
PR 22-AUG-1997; US-056632.
PR 22-AUG-1997; US-056636.
PR 22-AUG-1997; US-056637.
PR 22-AUG-1997; US-056662.
PR 22-AUG-1997; US-056664.
PR 22-AUG-1997; US-056845.
PR 22-AUG-1997; US-056862.
PR 22-AUG-1997; US-056864.
PR 22-AUG-1997; US-056872.
PR 22-AUG-1997; US-056874.
PR 22-AUG-1997; US-056875.
PR 22-AUG-1997; US-056876.
PR 22-AUG-1997; US-056877.
PR 22-AUG-1997; US-056878.
PR 22-AUG-1997; US-056879.
PR 22-AUG-1997; US-056880.
PR 22-AUG-1997; US-056881.
PR 22-AUG-1997; US-056882.
PR 22-AUG-1997; US-056884.
PR 22-AUG-1997; US-056886.
PR 22-AUG-1997; US-056887.
PR 22-AUG-1997; US-056888.
PR 22-AUG-1997; US-056889.
PR 22-AUG-1997; US-056892.
PR 22-AUG-1997; US-056893.
PR 22-AUG-1997; US-056894.
PR 22-AUG-1997; US-056903.
PR 22-AUG-1997; US-056908.
PR 22-AUG-1997; US-056909.
PR 22-AUG-1997; US-056910.
PR 22-AUG-1997; US-056911.
PR 05-SEP-1997; US-057650.
PR 05-SEP-1997; US-057669.
PR 05-SEP-1997; US-057761.
PR 12-SEP-1997; US-058785.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS,
PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
DR WPI; 98-506364/43.
DR P-PSDB; W74923.
PT New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 1; Page 434-435; 721pp; English.
CC This sequence represents a nucleic acid molecule designated Gene 14 from
CC the human cDNA clone HPMFD84 (deposited as clone ATCC 97897 and ATCC
CC 209043) which encodes a secreted human protein. The gene can be used to
CC generate fusion proteins by linking to the gene to a human immunoglobulin
CC Fc portion (e.g. V59502) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 186 novel genes and their fragments (nucleic
CC acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 186 polynucleotides, based on
CC which tissues they are most highly expressed in (see V59511 for described
CC uses).
SQ Sequence 1378 BP; 494 A; 254 C; 248 G; 377 T;

Query Match 1.9%; Score 37.8; DB 1; Length 1378;
Best Local Similarity 69.9%; Pred. No. 0.68;
Matches 51; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 1898 ctcgtatttatttttcgtacggttgtcgagctcaaaaataaaatcaaattaagctaaaaaa 1957
|| || |||||| || || || ! ||||| |||| ||| || |||||
Db 1263 CTTGTTCTTATTTTAATAAATTTATCAGAGTGAAAAAAAAAAAAAAAAAAAAAAAAAAA 1322

Qy 1958 aaaaaaaaaaaaa 1970
|||||||||||||
Db 1323 AAAAAAAAAAAAA 1335

RESULT 10
V59524
ID V59524 standard; DNA; 2323 BP.
AC V59524;
DT 06-JAN-1999 (first entry)
DE Human secreted protein gene 14 clone HPMFD84.
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS Homo sapiens.
PN WO9839448-A2.
PD 11-SEP-1998.
PF 06-MAR-1998; U04493.
PR 02-OCT-1997; US-061060.
PR 07-MAR-1997; US-038621.
PR 07-MAR-1997; US-040161.
PR 07-MAR-1997; US-040162.
PR 07-MAR-1997; US-040163.
PR 07-MAR-1997; US-040333.
PR 07-MAR-1997; US-040334.
PR 07-MAR-1997; US-040336.
PR 07-MAR-1997; US-040626.
PR 11-APR-1997; US-043311.
PR 11-APR-1997; US-043312.
PR 11-APR-1997; US-043313.
PR 11-APR-1997; US-043314.
PR 11-APR-1997; US-043568.
PR 11-APR-1997; US-043569.
PR 11-APR-1997; US-043576.
PR 11-APR-1997; US-043578.
PR 11-APR-1997; US-043580.
PR 11-APR-1997; US-043669.
PR 11-APR-1997; US-043670.
PR 11-APR-1997; US-043671.
PR 11-APR-1997; US-043672.

PR 11-APR-1997; US-043674.
PR 23-MAY-1997; US-047492.
PR 23-MAY-1997; US-047500.
PR 23-MAY-1997; US-047501.
PR 23-MAY-1997; US-047502.
PR 23-MAY-1997; US-047503.
PR 23-MAY-1997; US-047581.
PR 23-MAY-1997; US-047582.
PR 23-MAY-1997; US-047583.
PR 23-MAY-1997; US-047584.
PR 23-MAY-1997; US-047585.
PR 23-MAY-1997; US-047586.
PR 23-MAY-1997; US-047587.
PR 23-MAY-1997; US-047588.
PR 23-MAY-1997; US-047589.
PR 23-MAY-1997; US-047590.
PR 23-MAY-1997; US-047592.
PR 23-MAY-1997; US-047593.
PR 23-MAY-1997; US-047594.
PR 23-MAY-1997; US-047595.
PR 23-MAY-1997; US-047596.
PR 23-MAY-1997; US-047597.
PR 23-MAY-1997; US-047598.
PR 23-MAY-1997; US-047599.
PR 23-MAY-1997; US-047600.
PR 23-MAY-1997; US-047601.
PR 23-MAY-1997; US-047612.
PR 23-MAY-1997; US-047613.
PR 23-MAY-1997; US-047614.
PR 23-MAY-1997; US-047615.
PR 23-MAY-1997; US-047617.
PR 23-MAY-1997; US-047618.
PR 23-MAY-1997; US-047632.
PR 23-MAY-1997; US-047633.
PR 06-JUN-1997; US-048964.
PR 06-JUN-1997; US-048974.
PR 13-JUN-1997; US-049610.
PR 08-JUL-1997; US-051926.
PR 16-JUL-1997; US-052874.
PR 18-AUG-1997; US-055724.
PR 22-AUG-1997; US-056630.
PR 22-AUG-1997; US-056631.
PR 22-AUG-1997; US-056632.
PR 22-AUG-1997; US-056636.
PR 22-AUG-1997; US-056637.
PR 22-AUG-1997; US-056662.
PR 22-AUG-1997; US-056664.
PR 22-AUG-1997; US-056845.
PR 22-AUG-1997; US-056862.
PR 22-AUG-1997; US-056864.
PR 22-AUG-1997; US-056872.
PR 22-AUG-1997; US-056874.
PR 22-AUG-1997; US-056875.
PR 22-AUG-1997; US-056876.
PR 22-AUG-1997; US-056877.
PR 22-AUG-1997; US-056878.
PR 22-AUG-1997; US-056879.
PR 22-AUG-1997; US-056880.
PR 22-AUG-1997; US-056881.
PR 22-AUG-1997; US-056882.
PR 22-AUG-1997; US-056884.
PR 22-AUG-1997; US-056886.
PR 22-AUG-1997; US-056887.
PR 22-AUG-1997; US-056888.
PR 22-AUG-1997; US-056889.
PR 22-AUG-1997; US-056892.
PR 22-AUG-1997; US-056893.
PR 22-AUG-1997; US-056894.
PR 22-AUG-1997; US-056903.
PR 22-AUG-1997; US-056908.
PR 22-AUG-1997; US-056909.
PR 22-AUG-1997; US-056910.
PR 22-AUG-1997; US-056911.
PR 05-SEP-1997; US-057650.
PR 05-SEP-1997; US-057669.
PR 05-SEP-1997; US-057761.
PR 12-SEP-1997; US-058785.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS,
PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
DR WPI; 98-506364/43.
DR P-PSDB; W74744.
PT New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 1; Page 252-253; 721pp; English.
CC This sequence represents a nucleic acid molecule designated Gene 14 from
CC the human cDNA clone HPMFD84 (deposited as clone ATCC 97897 and ATCC
CC 209043) which encodes a secreted human protein. The gene can be used to
CC generate fusion proteins by linking to the gene to a human immunoglobulin
CC Fc portion (e.g. V59502) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 186 novel genes and their fragments (nucleic
CC acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 186 polynucleotides, based on
CC which tissues they are most highly expressed in (see V59511 for described
CC uses).
SQ Sequence 2323 BP; 760 A; 467 C; 438 G; 658 T;

Query Match 1.9%; Score 37.8; DB 1; Length 2323;
Best Local Similarity 69.9%; Pred. No. 0.87;
Matches 51; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 1898 ctcgtatttatttttcgtacggttgtcgagctcaaaaataaaatcaaattaagctaaaaaa 1957
|| || ||||||| || || || | ||||| |||| ||| || ||||||
Db 2201 CTTGTTCTTATTTTAATAAATTTATCAGAGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2260

Qy 1958 aaaaaaaaaaaaa 1970
|||||||||||||
Db 2261 AAAAAAAAAAAAA 2273

RESULT 11
Q72956
ID Q72956 standard; cDNA; 684 BP.
AC Q72956;
DT 04-JUN-1995 (first entry)
DE Clone pSP65-XaI-11 encoding recombinant Factor Xa inhibitor (FXaI).
KW Leech; factor Xa inhibitor; ss.
OS Hirudu medicinalis.
FH Key Location/Qualifiers
FT cds 139. .684
FT /*tag= a
FT misc_difference 1. .684
FT /*tag= b
FT /label= N= unspecified base
PN WO9423709-A.
PD 27-OCT-1994.
PF 08-APR-1994; U03918.
PR 09-APR-1993; US-045804.
PA (BIOT-) BIO-TECHNOLOGY GENERAL CORP.
PA (YISS) YISSUM RES & DEV CO.
PI Fischer M, Goldlust A, Guy R, Levanon A, Panet A;
PI Rigbi M, Werber MM, Zeelon EP;
DR WPI; 94-341457/42.
DR P-PSDB; R62619.
PT Recombinant factor 10a inhibitor of Hirudo medicinalis - for
PT treating excessive blood coagulation, partic. thrombosis, also
PT related DNA, vectors, transformed cells and antibodies

```
PS   Claim 3; Figure 10; 107pp; English.
CC   Total RNA was extracted from 120 leeches and used to prepare a
CC   cDNA library which was screened for clones contg. FXaI DNA using
CC   as probe radiolabeled DNA from the PCR derived plasmid pSP65-XaI-
CC   4 (see Q72957). The plasmid of one of the positive clones - pSK-
CC   XaI-11 (clone 11) contains a 684 bp insert (see Q72956) essentially
CC   identical to the previously obtd. PCR derived sequences but contg.
CC   78 additional bps at its 5' end. It encodes a prepeptide of 110
CC   AAs (incl. the N-terminal Met) extending from Met 1 to Gly 110 (see
CC   R62619). AAs 1-25 constitute a leader sequence, and the mature
CC   protein comprises Tyr 26 to Gly 110. The sequnce Tyr 2 - Gly 86
CC   of clone 4 (Q72957) is equiv. to the sequence Tyr 26 - Gly 110
CC   of clone 11 (Q72956). There are only two differences between
CC   clones 4 and 11. The polypeptide expressed by clone 11 is
CC   referred to as recombinant FXaI. A plasmid comprising the DNA
CC   designated pSP65-XaI-11 and deposited under ATCC Accession No.
CC   69138 is claimed.
SQ   Sequence   684 BP;    263 A;    110 C;    120 G;    190 T;


  Query Match             1.9%;  Score 37;  DB 1;  Length 684;
  Best Local Similarity   71.0%;  Pred. No. 0.8;
  Matches   49;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy   1902 tatttatttttcgtacggttgtcgagctcaaaaataaaatcaaattaagctaaaaaaaaaa 1961
            ||| | || |||| ||  | || |  |  |||| ||||  |||  ||   ||||||||||
Db    609 TATATTTTATCGTTCGCATATCAACATTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 668

Qy   1962 aaaaaaaaa 1970
          |||||||||
Db    669 AAAAAAAAA 677


RESULT  12
V09108
ID   V09108 standard; cDNA; 314 BP.
AC   V09108;
DT   28-MAY-1998  (first entry)
DE   3' nucleotide sequence of clone G55_1.
KW   G55_1 clone; cDNA library; cell proliferation; cell differentiation;
KW   ss.
OS   Homo sapiens.
PN   US5708157-A.
PD   13-JAN-1998.
PF   26-JUL-1996; 686878.
PR   26-JUL-1996; US-686878.
PA   (GEMY ) GENETICS INST INC.
PA   (JACO/) JACOBS K.
PI   Evans C, Jacobs K, LaVallie ER, McCoy JM, Merberg D,
PI   Racie LA, Spaulding V, Treacy M;
DR   WPI; 98-109352/10.
PT   Human testes derived cDNA and protein(s) - may be useful for
PT   laboratory use, or for stimulating cell proliferation and
PT   differentiation
PS   Disclosure; Column 63-64; 48pp; English.
CC   This sequence is the 3' nucleotide sequence for the G55_1 clone, which
CC   is isolated from a human PBMC cDNA library.  When the G55_1 sequence
CC   was searched in the Genbank database, the clone demonstrated at least
CC   some identity with an EST identified as "yp16a07.rl Homo sapiens cDNA
CC   clone 187572 5'".
SQ   Sequence   314 BP;    110 A;    67 C;    50 G;    64 T;


  Query Match             1.9%;  Score 36.6;  DB 1;  Length 314;
  Best Local Similarity   62.1%;  Pred. No. 0.71;
  Matches   54;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy   1884 gtccttagtcttaactcgtatttatttcgtacggttgtcgagctcaaaaataaaatcaa 1943
          | ||||| | | ||   ||    |||   | | || ||  || |||  ||||  ||
Db    227 GGCCTTANTTTAAAAAATTAAANTTTTNANAAAGTTTTTCTTACTAAAAGGGAAAAAAAA 286

Qy   1944 attaagctaaaaaaaaaaaaaaaaaa 1970
          |  ||  ||||||||||||||||||||
Db    287 AAAAAAAAAAAAAAAAAAAAAAAAA 313


RESULT  13
Q20263
ID   Q20263 standard; cDNA; 1118 BP.
AC   Q20263;
DT   25-MAR-1992  (first entry)
DE   Deg-1 gene.
KW   Neurodegeneration; cell death; nematode; Huntington's; Lou Gehrig's;
KW   Alzheimer's; roundworm; ss.
OS   Caenorhabditis elegans.
FH   Key             Location/Qualifiers
FT   cds             3. .887
FT                   /*tag=  a
FT   polya_site      1028. .1118
FT                   /*tag= b
PN   WO9119007-A.
PD   12-DEC-1991.
PF   30-MAY-1991; U03826.
PR   30-MAY-1990; US-530986.
PA   (UYCO-) COLUMIBA UNIV NEW Y.
PI   Chalfie M, Wolinsky E, Driscoll M;
DR   WPI; 92-007488/01.
DR   P-PSDB; R20106.
PT   DNA sequences involved in neuronal degeneration - used for
PT   detecting, diagnosing, treating and preventing degenerative
PT   diseases.
PS   Claim 10; Fig 7; 57pp; English.
CC   The deg-1 (degradation) gene was cloned by transposon tagging.  Two
CC   partial deg-1 cDNAs were isolated from a lambda gt10 library prepd.
CC   from cDNA from 12-hour old larvae.  The larger cDNA, given here,
CC   comprises an open reading frame of 884 bp followed by 141 bp and a
CC   poly A tail. The smaller cDNA lacks a 72 bp sequence from within
CC   the ORF (position not specified).  Since no untranslated regions
CC   appear to precede the ORFs , the cDNAs seem to be incomplete.  The
CC   two sequences may represent alternatively spliced variants.  The
CC   deg-1 gene appears to encode a membrane protein, probably a recep-
CC   tor.  Mutated versions of this protein are associated with the
CC   late-onset degeneration of a small set of neurons.  Strains of C.
CC   elegans having mutations in the deg-1 gene (u38 and uIn1) have been
CC   identified in a screen for touch sensitivity.  The DNA may be use-
CC   ful for detecting, diagnosing, preventing and treating degenerative
CC   diseases such as Alzheimer's disease, amyotrophic lateral sclerosis
CC   and Huntington's disease.
CC   See also Q20264.
SQ   Sequence   1118 BP;    395 A;    195 C;    222 G;    306 T;


  Query Match             1.9%;  Score 36.6;  DB 1;  Length 1118;
  Best Local Similarity   65.1%;  Pred. No. 1.3;
  Matches   54;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy   1888 ttagtcttaactcgtatttatttttcgtacggttgtcgagctcaaaaataaaatcaaatta 1947
          ||  ||  | |||  |||   | | ||  | || | ||    ||||| ||||  |||  |
Db    989 TTTTTCCGATCTCTAATTGTATATAGTGAGCTTTTTGATTAAAAAAAAAAAAAAAAAAAA 1048

Qy   1948 agctaaaaaaaaaaaaaaaaaa 1970
          |   ||||||||||||||||||
Db   1049 AAAAAAAAAAAAAAAAAAAAAA 1071


RESULT  14
Q35987
ID   Q35987 standard; cDNA; 2116 BP.
AC   Q35987;
DT   18-JUN-1993  (first entry)
DE   Tomato hsp80 cDNA sequence.
KW   Tomato; heat shock protein; 80.5; hsp80; prokaryotic; eukaryotic;
KW   locus A; locus B; copy number; promoter; meristematic tissue;
KW   flower meristems; fruit; root; shoot; apices; floral primordia; ss.
OS   Lycopersicon esculentum.
FH   Key             Location/Qualifiers
```

```
FT   cds             3. .1865
FT                   /*tag=  a
PN   US5187267-A.
PD   16-FEB-1993.
PF   19-JUN-1990; 541883.
PR   19-JUN-1990; US-541883.
PA   (CALJ ) CALGENE INC.
PI   Comai L,  Koning AJ;
DR   WPI; 93-075789/09.
DR   P-PSDB; R32548.
PT   New DNA constructs - comprising tomato heat shock protein 80 gene
PT   non-coding region and gene region of interest, useful for
PT   producing transgenic plants with modified phenotype(s)
PS   Claim 2; Fig 1; 33pp; English.
CC   The sequence given encodes the tomato heat shock protein 80.5 (hsp80).
CC   Tomato hsp80 is characterised by having a mRNA of about 2.3 kb, two
CC   introns, a predicted pI of about 4.69 and a molecular weight of
CC   80,479.8 daltons.  The polypeptides encoded by tomato hsp80 have
CC   substantial homology with the polypeptides encoded by related heat
CC   shock proteins from a variety of prokaryotic and eukaryotic species.
CC   There are two hsp80 genes in tomato, identified as locus A and locus
CC   B.  The low copy number of the hsp80 gene is a useful indicator of the
CC   strength of the promoter.  Tomato hsp80 message is found in abundance
CC   in most, if not all meristematic tissue eg, flower meristems, early
CC   fruit, and root and shoot apices.  hsp80 message is detected in floral
CC   primordia of immature fruit, where it increases until the fruit begins
CC   to swell.  The mRNA levels observed in normal mature leaves is approx.
CC   1/10th of the level observed in meristematic tissue.
SQ   Sequence   2116 BP;    647 A;    349 C;    562 G;    558 T;


  Query Match             1.9%;  Score 36.6;  DB 1;  Length 2116;
  Best Local Similarity   60.6%;  Pred. No. 1.8;
  Matches   60;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy   1872 ctttttgtcatagtccttagtcttaactcgtatttatttttcgtacggttgtcgagctcaa 1931
            ||||||    | | | || ||  |   |    ||||||||| | |        |    ||
Db   2007 CTTTTTTGTTTTGACATTCGTGGTTGATATAGTTTATTTTCTTGCCCAAAAAAAAAAAAA 2066

Qy   1932 aaataaaatcaaattaagctaaaaaaaaaaaaaaaaaa 1970
               ||| ||||  |||  ||   ||||||||||||||||||
Db   2067 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2105


RESULT  15
N60848
ID   N60848 standard; DNA; 3628 BP.
AC   N60848;
DT   24-OCT-1991  (first entry)
DE   Plasmid sequence encoding bovine pre-prolactin hormone.
KW   E.coli; growth hormone; bGH; bPRL; ds.
OS   Bos taurus.
PN   J61202691-A.
PD   08-SEP-1986.
PF   04-MAR-1985; 042406.
PR   04-MAR-1985; JP-042406.
PA   (NAKA/) NAKAJIMA K.
DR   WPI; 86-276389/42.
PT   Plasmid(s) contg. ampicillin resistant genes - for amplification
PT   of bovine growth hormone and prolactin genes.
PS   Disclosure; Table 1-18; 13pp; Japanese.
CC   The bovine pre-prolactin hormone encoding sequence is derived from
CC   pituitary tissue. The plasmids carry an amp resistance gene and,
CC   and may be expressed by an E.coli host in an ampicillin medium
CC   for efficient and large scale production of the hormone.
SQ   Sequence   3628 BP;    965 A;    873 C;    901 G;    889 T;


  Query Match             1.9%;  Score 36.6;  DB 1;  Length 3628;
  Best Local Similarity   65.1%;  Pred. No. 2.3;
  Matches   54;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy   1888 ttagtcttaactcgtatttattttcgtacggttgtcgagctcaaaaataaaatcaaatta 1947
            ||  ||||||  |   |      |  |||  | ||||  |  || ||||  ||||  |||  |
Db   1183 TTCATCTTAAATAAAAACAGACTCTGTAGCGATGTCAAAATCTAAAAAAAAAAAAAAAAA 1242

Qy   1948 agctaaaaaaaaaaaaaaaaaa 1970
               |   ||||||||||||||||||
Db   1243 AAAAAAAAAAAAAAAAAAAAAA 1265


Search completed: September  4, 2000, 00:39:04
Job time: 16919 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2000, 23:42:35 ; Search time 11551.9 Seconds
(without alignments)
304.508 Million cell updates/sec

Title: US-09-332-522B-3
Perfect score: 1971
Sequence: 1 cggcacgaggattaatgctg.....aaaaaaaaaaaaaaaaaaac 1971

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: em_fun:*
17: em_hum1:*
18: em_hum2:*
19: em_in:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_sy:*
29: em_un:*
30: em_vi:*
31: gb_htg1:*
32: gb_htg2:*
33: gb_in1:*
34: gb_in2:*
35: em_ba1:*
36: em_ba2:*
37: em_hum3:*
38: em_hum4:*
39: gb_pr4:*
40: gb_htg3:*
41: gb_htg4:*
42: gb_htg5:*
43: gb_htg6:*
44: gb_htg7:*
45: em_htg1:*
46: em_htg2:*
47: em_htg3:*
48: em_hum5:*
49: gb_pl3:*
50: gb_pr5:*
51: gb_htg8:*
52: gb_htg9:*
53: gb_htg10:*
54: gb_htg11:*
55: gb_htg12:*
56: gb_htg13:*
57: gb_htg14:*
58: gb_in3:*
59: gb_htg15:*
60: gb_htg16:*
61: gb_htg17:*
62: em_htg4:*
63: em_htg5:*
64: em_htg6:*
65: em_htg7:*
66: em_hum6:*
67: gb_htg18:*
68: gb_htg19:*
69: gb_htg20:*
70: gb_htg21:*
71: gb_htg22:*
72: gb_htg23:*
73: gb_htg24:*
74: gb_htg25:*
75: gb_htg26:*
76: gb_htg27:*
77: gb_htg28:*
78: gb_htg29:*
79: gb_htg30:*
80: gb_htg31:*
81: gb_vi1:*
82: gb_vi2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1806.4	91.6	77862	34	AC005465	AC005465 Drosophil
	2	1806.4	91.6	78927	43	AC020317	AC020317 Drosophil
	3	1806.4	91.6	249001	58	AE003824	AE003824 Drosophil
	4	88.6	4.5	1732	12	AF019611	AF019611 Cricetulu
	5	86	4.4	1759	11	AF019612	AF019612 Homo sapi
c	6	46.4	2.4	3946	33	DDU25143	U25143 Dictyosteli
	7	46	2.3	7218	5	I66494	I66494 Sequence 14
	8	45	2.3	108908	33	PFMAL3P8	AL034560 Plasmodiu
	9	44.6	2.3	173184	60	AC013583	AC013583 Homo sapi
c	10	44.6	2.3	237302	40	AL136310	AL136310 Homo sapi
	11	44.2	2.2	268	33	DMTNF14B	X01937 Drosophila
c	12	44.2	2.2	17196	34	AE003083	AE003083 Drosophil
c	13	44.2	2.2	24781	34	AE003098	AE003098 Drosophil
c	14	44.2	2.2	51236	41	AC006589	AC006589 Drosophil
c	15	44.2	2.2	61267	34	AC004336	AC004336 Drosophil
	16	44.2	2.2	82692	19	AC003924	Ac003924 Drosophil
c	17	44.2	2.2	100729	41	AC008141	AC008141 Drosophil
c	18	44.2	2.2	104494	51	AC012166	AC012166 Drosophil
	19	44.2	2.2	129241	51	AC010113	AC010113 Drosophil
	20	44.2	2.2	139238	19	AC006302	Ac006302 Drosophil
c	21	44.2	2.2	139561	55	AC007904	AC007904 Drosophil
	22	44.2	2.2	174982	54	AC013424	AC013424 Drosophil
	23	44.2	2.2	176878	42	AC006091	AC006091 Drosophil
	24	44.2	2.2	198282	61	AC007415	AC007415 Drosophil

	25	44.2	2.2 224263	34	AE003720	AE003720 Drosophil
c	26	44.2	2.2 224263	34	AE003720	AE003720 Drosophil
	27	44.2	2.2 225419	34	AE003703	AE003703 Drosophil
c	28	44.2	2.2 229655	34	AE003707	AE003707 Drosophil
c	29	44.2	2.2 230977	34	AE003729	AE003729 Drosophil
	30	44.2	2.2 251154	34	AE003656	AE003656 Drosophil
	31	44.2	2.2 253172	34	AE003718	AE003718 Drosophil
	32	44.2	2.2 254846	34	AE003725	AE003725 Drosophil
	33	44.2	2.2 257672	34	AE003517	AE003517 Drosophil
c	34	44.2	2.2 261000	34	AE003654	AE003654 Drosophil
c	35	44.2	2.2 267271	58	AE003800	AE003800 Drosophil
	36	44.2	2.2 303885	34	DROSADH01	AE003407 Drosophil
c	37	44.2	2.2 316735	58	AE003786	AE003786 Drosophil
	38	43.2	2.2 93240	42	AC007586	AC007586 Drosophil
c	39	43.2	2.2 134837	55	AC007888	AC007888 Drosophil
c	40	43.2	2.2 179122	34	AC006414	AC006414 Drosophil
c	41	43.2	2.2 224939	34	AE003710	AE003710 Drosophil
	42	42.6	2.2 1546	34	AE003035	AE003035 Drosophil
c	43	42.6	2.2 40653	51	AC011906	AC011906 Drosophil
c	44	42.6	2.2 122539	51	AC010110	AC010110 Drosophil
	45	42.6	2.2 127091	53	AC009843	AC009843 Drosophil

ALIGNMENTS

RESULT 1
AC005465
LOCUS AC005465 77862 bp DNA INV 19-NOV-1998
DEFINITION Drosophila melanogaster, chromosome 2R, region 48C1-48C2, P1 clone DS05372, complete sequence.
ACCESSION AC005465
VERSION AC005465.1 GI:3893020
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 77862)
AUTHORS Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Punch,E., Sequeira,A., Sethi,H., Snir,E., Svirskas,R.R., Twomey,B., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.
TITLE Sequencing of Drosophila chromosome 2R, region 48C1-48C2
JOURNAL Unpublished (1998)
REFERENCE 2 (bases 1 to 77862)
AUTHORS Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.E.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-1998) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT On Nov 19, 1998 this sequence version replaced gi:3746016.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://www.fruitfly.org/sequence/) or send email to bdgp@fruitfly.berkeley.edu.
P1 library location: 56-92.

FEATURES Location/Qualifiers
source 1. .77862
/organism="Drosophila melanogaster"
/strain="y2; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2R"
/map="48C1-48C2"
/clone_lib="P1 library, partial Sau3A in pAd10sacBII"
/clone="P1 DS05372 (D379)"
/note="Between bp 20,516 and bp 20,517, Tn10 and the 9 bp duplication were excised."
BASE COUNT 21405 a 16792 c 16981 g 22684 t
ORIGIN

Query Match 91.6%; Score 1806.4; DB 34; Length 77862;
Best Local Similarity 94.2%; Pred. No. 0;
Matches 1945; Conservative 0; Mismatches 1; Indels 118; Gaps 2;

```
Qy      10 gattaatgctgatttctggtctggactacacagcattgctggtataaggagtcgggacca 69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13013 GATTAATGCTGATTTCTGGTCTGGACTACACAGCATTGCTGGTATAAGGAGTCGGGACCA 13072

Qy      70 gaggagtaagatttcgggaaggaatcccgtccggtagggactactagcattcgcaagtga 129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13073 GAGGAGTAAGATTTCGGGAAGGAATCCCGTCCGGTAGGGACTACTAGCATTCGCAAGTGA 13132

Qy     130 cgtccagcaaccggaggacccccaactgtagaatccgcatcaccatcctaatcccaacaa 189
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13133 CGTCCAGCAACCGGAGGACCCCCAACTGTAGAATCCGCATCACCATCCTAATCCCAACAA 13192

Qy     190 accaatgacatcttgagacctcaccagccatggatcccttcgtgttcttcatagtactgg 249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13193 ACCAATGACATCTTGAGACCTCACCAGCCATGGATCCCTTCGTGTTCTTCATAGTACTGG 13252

Qy     250 catcgctttatggcgttctttacttttttcgaccgcttcttca---------------- 291
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13253 CATCGCTTTATGGCGTTCTTTACTTTTTCGACCGCTTCTTCAAGGTGTAGTATATCCAGC 13312

Qy     292 --------------------------------agagttgcatgcactacccgtacgatg 318
                                           ||||||||||||||||||||||||||||
Db   13313 CAAAGTACGTCCAGATACTTAATGTAATCCCTTAGAGTTGCATGCACTACCCGTACGATG 13372

Qy     319 ccttcctcaagaacaccgggctgagtataaatttcatgagcctccactggcacacgagtg 378
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13373 CCTTCCTCAAGAACACCGGGCTGAGTATAAATTTCATGAGCCTCCACTGGCACACGAGTG 13432

Qy     379 cctttaacaggaccctcctacgctggggatctgccggtaacagctgcacccggagagtaa 438
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13433 CCTTTAACAGGACCCTCCTACGCTGGGGATCTGCCGGTAACAGCTGCACCCGGAGAGTAA 13492

Qy     439 tgatcaccagctttaatgtaggagtcctggtcacctttttctctgctcccgatcggtctga 498
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13493 TGATCACCAGCTTTAATGTAGGAGTCCTGGTCACCTTTTCTCTGCTCCCGATCGGTCTGA 13552

Qy     499 tcctgctcattgccactatcttcagcagtggtgaacaagacagctcttcgtctgtatcct 558
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13553 TCCTGCTCATTGCCACTATCTTCAGCAGTGGTGAACAAGACAGCTCTTCGTCTGTATCCT 13612

Qy     559 cgcccgttggagtccctgtgcagctggaaattctactgcccggcgtcaacttgccgttgg 618
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13613 CGCCCGTTGGAGTCCCTGTGCAGCTGGAAATTCTACTGCCCGGCGTCAACTTGCCGTTGG 13672

Qy     619 aggagatcggatactacatcacaacccttgtgctctgcttggtggtgcacgagatgggac 678
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13673 AGGAGATCGGATACTACATCACAACCCTTGTGCTCTGCTTGGTGGTGCACGAGATGGGAC 13732

Qy     679 acgccctggccgctgtgatggaggatgtgcctgtcaccgggtttggaataaa--------- 729
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13733 ACGCCCTGGCCGCTGTGATGGAGGATGTGCCTGTCACCGGGTTTGGAATAAAGGTTAGAA 13792

Qy     730 ------------------------------------------------------------ag 731
                                                                        ||
```

```
Db  13793 TGATATAGCGTACTTTACAACCTAATACGCTTACTAATTAGGTAAATATTAATATTTCAG 13852

Qy    732 ttcatcttctgcctgccgttagcatacacggagctctcccacgaccacttaaacagtcta 791
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13853 TTCATCTTCTGCCTGCCGTTAGCATACACGGAGCTCTCCCACGACCACTTAAACAGTCTA 13912

Qy    792 cgttggttccgcaagctacgtgttctgtgcgctggaatctggcataattttgtgttcgct 851
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13913 CGTTGGTTCCGCAAGCTACGTGTTCTGTGCGCTGGAATCTGGCATAATTTTGTGTTCGCT 13972

Qy    852 ggcgtgtgctatctcttaatctcaacggtgggaatcactatgtcacctttgtacgcttac 911
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13973 GGCGTGTGCTATCTCTTAATCTCAACGGTGGGAATCACTATGTCACCTTTGTACGCTTAC 14032

Qy    912 aaccaacacgtagtggtcactgaactaacaaggaaatccccgctgaggggagagcgcggc 971
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14033 AACCAACACGTAGTGGTCACTGAACTAACAAGGAAATCCCCGCTGAGGGGAGAGCGCGGC 14092

Qy    972 ttgcaagtggacaatcaaataacccaagtaaacggctgcccagtaaacagcgaggagagt 1031
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14093 TTGCAAGTGGACAATCAAATAACCCAAGTAAACGGCTGCCCAGTAAACAGCGAGGAGAGT 14152

Qy   1032 tgggtgacatgcctgcagaactctctgaagctcaagccgggctactgtgtgagtgcggac 1091
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14153 TGGGTGACATGCCTGCAGAACTCTCTGAAGCTCAAGCCGGGCTACTGTGTGAGTGCGGAC 14212

Qy   1092 ttcgtgcagcttaacgacgaaagcagcgccatctcacatcatagcattgatggtcagcta 1151
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14213 TTCGTGCAGCTTAACGACGAAAGCAGCGCCATCTCACATCATAGCATTGATGGTCAGCTA 14272

Qy   1152 cagtgctgtgatgaactaaatccgaacgtaagctgcttcgaggtggtggaggacgcaaat 1211
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14273 CAGTGCTGTGATGAACTAAATCCGAACGTAAGCTGCTTCGAGGTGGTGGAGGACGCAAAT 14332

Qy   1212 ggagatgtgccggtggagctgccgcagcatgtatgtctcaatgtgcgccgcactttggag 1271
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14333 GGAGATGTGCCGGTGGAGCTGCCGCAGCATGTATGTCTCAATGTGCGCCGCACTTTGGAG 14392

Qy   1272 gaggtctccgagcactgctcgtccggagtttgcaacgagggattctgcctacgaccgctt 1331
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14393 GAGGTCTCCGAGCACTGCTCGTCCGGAGTTTGCAACGAGGGATTCTGCCTACGACCGCTT 14452

Qy   1332 atacgaaatatcactgccataatgacgttcaagcgacagaattttcgcggagagaagctg 1391
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14453 ATACGAAATATCACTGCCATAATGACGTTCAAGCGACAGAATTTTCGCGGAGAGAAGCTG 14512

Qy   1392 ccgccggtgatctatgtgggccatccatgggatgtcactcgaactgtggaggtatccgcc 1451
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14513 CCGCCGGTGATCTATGTGGGCCATCCATGGGATGTCACTCGAACTGTGGAGGTATCCGCC 14572

Qy   1452 tttgtgccgagatatagcttattaaaggcagcctggccggatgcctggctgctgctcctc 1511
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14573 TTTGTGCCGAGATATAGCTTATTAAAGGCAGCCTGGCCGGATGCCTGGCTGCTGCTCCTC 14632

Qy   1512 aagtataacgtggtcttcagcataggattggcgttgatcaatgccattccctgctttggt 1571
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14633 AAGTATAACGTGGTCTTCAGCATAGGATTGGCGTTGATCAATGCCATTCCCTGCTTTGGT 14692

Qy   1572 ttcgatggcgcccacattaccagcaccgtgatacacagcttcttggtgggcagagtggat 1631
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14693 TTCGATGGCGCCCACATTACCAGCACCGTGATACACAGCTTCTTGGTGGGCAGAGTGGAT 14752

Qy   1632 cagcatgccaagagagatatcatctcgttgataatcaccagcgtgggttcccttctcttt 1691
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14753 CAGCATGCCAAGAGAGATATCATCTCGTTGATAATCACCAGCGTGGGTTCCCTTCTCTTT 14812

Qy   1692 gcactggccctgcttaaggtggcctggttgagttttctgcgacccctgctttaagaactg 1751
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14813 GCACTGGCCCTGCTTAAGGTGGCCTGGTTGAGTTTTCTGCGACCCCTGCTTTAAGAACTG 14872

Qy   1752 aaatggaaaactgaaatggatcctgggagttcaactccctgcaaagacgctagactgcta 1811
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14873 AAATGGAAAACTGAAATGGATCCTGGGAGTTCAACTCCCTGCAAAGACGCTAGACTGCTA 14932
```

```
Qy   1812 tttcaccttcacgaaacacacaaaaacacagcgaattgtagcacctcaaagattcgatag 1871
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14933 TTTCACCTTCACGAAACACACAAAAACACAGCGAATTGTAGCACCTCAAAGATTCGATAG 14992

Qy   1872 ctttttgtcatagtccttagtcttaactcgtatttatttttcgtacggttgtcgagctcaa 1931
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14993 CTTTTTGTCATAGTCCTTAGTCTTAACTCGTATTTATTTTCGTACGGTTGTCGAGCTCAA 15052

Qy   1932 aaataaaatcaaattaagctaaaa 1955
          ||||||||||||||||||||||| |
Db  15053 AAATAAAATCAAATTAAGCTAAGA 15076


RESULT   2
AC020317
LOCUS       AC020317    78927 bp    DNA             HTG       03-JAN-2000
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC020317
VERSION     AC020317.1  GI:6664580
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 78927)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212749 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .78927
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    22005 a  17623 c  17956 g  21343 t
ORIGIN


  Query Match             91.6%;  Score 1806.4;  DB 43;  Length 78927;
  Best Local Similarity   94.2%;  Pred. No. 0;
  Matches 1945;  Conservative    0;  Mismatches    1;  Indels  118;  Gaps    2;

Qy     10 gattaatgctgatttctggtctggactacacagcattgctggtataaggagtcgggacca 69
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  52202 GATTAATGCTGATTTCTGGTCTGGACTACACAGCATTGCTGGTATAAGGAGTCGGGACCA 52261

Qy     70 gaggagtaagatttcgggaaggaatcccgtccggtagggactactagcattcgcaagtga 129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  52262 GAGGAGTAAGATTTCGGGAAGGAATCCCGTCCGGTAGGGACTACTAGCATTCGCAAGTGA 52321

Qy    130 cgtccagcaaccggaggaccccccaactgtagaatccgcatcaccatcctaatcccaacaa 189
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  52322 CGTCCAGCAACCGGAGGACCCCCAACTGTAGAATCCGCATCACCATCCTAATCCCAACAA 52381

Qy    190 accaatgacatcttgagacctcaccagccatggatcccttcgtgttcttcatagtactgg 249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  52382 ACCAATGACATCTTGAGACCTCACCAGCCATGGATCCCTTCGTGTTCTTCATAGTACTGG 52441

Qy    250 catcgctttatggcgttctttacttttttcgaccgcttcttca---------------- 291
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  52442 CATCGCTTTATGGCGTTCTTTACTTTTTCGACCGCTTCTTCAAGGTGTAGTATATCCAGC 52501

Qy    292 ---------------------------------agagttgcatgcactacccgtacgatg 318
                                            ||||||||||||||||||||||||||
Db  52502 CAAAGTACGTCCAGATACTTAATGTAATCCCTTAGAGTTGCATGCACTACCCGTACGATG 52561
```

```
Qy    319 ccttcctcaagaacaccgggctgagtataaaatttcatgagcctccactggcacacgagtg 378
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  52562 CCTTCCTCAAGAACACCGGGCTGAGTATAAATTTCATGAGCCTCCACTGGCACACGAGTG 52621

Qy    379 cctttaacaggaccctcctacgctggggatctgccggtaacagctgcacccggagagtaa 438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  52622 CCTTTAACAGGACCCTCCTACGCTGGGGATCTGCCGGTAACAGCTGCACCCGGAGAGTAA 52681

Qy    439 tgatcaccagctttaatgtaggagtcctggtcacctttttctctgctcccgatcggtctga 498
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  52682 TGATCACCAGCTTTAATGTAGGAGTCCTGGTCACCTTTTCTCTGCTCCCGATCGGTCTGA 52741

Qy    499 tcctgctcattgccactatcttcagcagtggtgaacaagacagctcttcgtctgtatcct 558
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  52742 TCCTGCTCATTGCCACTATCTTCAGCAGTGGTGAACAAGACAGCTCTTCGTCTGTATCCT 52801

Qy    559 cgcccgttggagtccctgtgcagctggaaattctactgcccggcgtcaacttgccgttgg 618
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  52802 CGCCCGTTGGAGTCCCTGTGCAGCTGGAAATTCTACTGCCCGGCGTCAACTTGCCGTTGG 52861

Qy    619 aggagatcggatactacatcacaacccttgtgctctgcttggtggtgcacgagatgggac 678
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  52862 AGGAGATCGGATACTACATCACAACCCTTGTGCTCTGCTTGGTGGTGCACGAGATGGGAC 52921

Qy    679 acgccctggccgctgtgatggaggatgtgcctgtcaccgggtttggaataa--------- 729
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  52922 ACGCCCTGGCCGCTGTGATGGAGGATGTGCCTGTCACCGGGTTTGGAATAAAGGTTAGAA 52981

Qy    730 ----------------------------------------------------------ag 731
                                                                       ||
Db  52982 TGATATAGCGTACTTTACAACCTAATACGCTTACTAATTAGGTAAATATTAATATTTCAG 53041

Qy    732 ttcatcttctgcctgccgttagcatacacggagctctcccacgaccacttaaacagtcta 791
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53042 TTCATCTTCTGCCTGCCGTTAGCATACACGGAGCTCTCCCACGACCACTTAAACAGTCTA 53101

Qy    792 cgttggttccgcaagctacgtgttctgtgcgctggaatctggcataattttgtgttcgct 851
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53102 CGTTGGTTCCGCAAGCTACGTGTTCTGTGCGCTGGAATCTGGCATAATTTTGTGTTCGCT 53161

Qy    852 ggcgtgtgctatctcttaatctcaacggtgggaatcactatgtcacctttgtacgcttac 911
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53162 GGCGTGTGCTATCTCTTAATCTCAACGGTGGGAATCACTATGTCACCTTTGTACGCTTAC 53221

Qy    912 aaccaacacgtagtggtcactgaactaacaaggaaatccccgctgaggggagagcgcggc 971
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53222 AACCAACACGTAGTGGTCACTGAACTAACAAGGAAATCCCCGCTGAGGGGAGAGCGCGGC 53281

Qy    972 ttgcaagtggacaatcaaataacccaagtaaacggctgcccagtaaacagcgaggagagt 1031
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53282 TTGCAAGTGGACAATCAAATAACCCAAGTAAACGGCTGCCCAGTAAACAGCGAGGAGAGT 53341

Qy   1032 tgggtgacatgcctgcagaactctctgaagctcaagccgggctactgtgtgagtgcggac 1091
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53342 TGGGTGACATGCCTGCAGAACTCTCTGAAGCTCAAGCCGGGCTACTGTGTGAGTGCGGAC 53401

Qy   1092 ttcgtgcagcttaacgacgaaagcagcgccatctcacatcatagcattgatggtcagcta 1151
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53402 TTCGTGCAGCTTAACGACGAAAGCAGCGCCATCTCACATCATAGCATTGATGGTCAGCTA 53461

Qy   1152 cagtgctgtgatgaactaaatccgaacgtaagctgcttcgaggtggtggaggacgcaaat 1211
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53462 CAGTGCTGTGATGAACTAAATCCGAACGTAAGCTGCTTCGAGGTGGTGGAGGACGCAAAT 53521

Qy   1212 ggagatgtgccggtggagctgccgcagcatgtatgtctcaatgtgcgccgcactttggag 1271
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53522 GGAGATGTGCCGGTGGAGCTGCCGCAGCATGTATGTCTCAATGTGCGCCGCACTTTGGAG 53581

Qy   1272 gaggtctccgagcactgctcgtccggagtttgcaacgagggattctgcctacgaccgctt 1331
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53582 GAGGTCTCCGAGCACTGCTCGTCCGGAGTTTGCAACGAGGGATTCTGCCTACGACCGCTT 53641
```

```
Qy   1332 atacgaaatatcactgccataatgacgttcaagcgacagaattttcgcggagagaagctg 1391
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53642 ATACGAAATATCACTGCCATAATGACGTTCAAGCGACAGAATTTTCGCGGAGAGAAGCTG 53701

Qy   1392 ccgccggtgatctatgtgggccatccatgggatgtcactcgaactgtggaggtatccgcc 1451
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53702 CCGCCGGTGATCTATGTGGGCCATCCATGGGATGTCACTCGAACTGTGGAGGTATCCGCC 53761

Qy   1452 tttgtgccgagatatagcttattaaaggcagcctggccggatgcctggctgctgctcctc 1511
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53762 TTTGTGCCGAGATATAGCTTATTAAAGGCAGCCTGGCCGGATGCCTGGCTGCTGCTCCTC 53821

Qy   1512 aagtataacgtggtcttcagcataggattggcgttgatcaatgccattccctgctttggt 1571
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53822 AAGTATAACGTGGTCTTCAGCATAGGATTGGCGTTGATCAATGCCATTCCCTGCTTTGGT 53881

Qy   1572 ttcgatggcgcccacattaccagcaccgtgatacacagcttcttggtgggcagagtggat 1631
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53882 TTCGATGGCGCCCACATTACCAGCACCGTGATACACAGCTTCTTGGTGGGCAGAGTGGAT 53941

Qy   1632 cagcatgccaagagagatatcatctcgttgataatcaccagcgtgggttccctttctcttt 1691
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53942 CAGCATGCCAAGAGAGATATCATCTCGTTGATAATCACCAGCGTGGGTTCCCTTCTCTTT 54001

Qy   1692 gcactggccctgcttaaggtggcctggttgagttttctgcgacccctgctttaagaactg 1751
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54002 GCACTGGCCCTGCTTAAGGTGGCCTGGTTGAGTTTTCTGCGACCCCTGCTTTAAGAACTG 54061

Qy   1752 aaatggaaaactgaaatggatcctgggagttcaactccctgcaaagacgctagactgcta 1811
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54062 AAATGGAAAACTGAAATGGATCCTGGGAGTTCAACTCCCTGCAAAGACGCTAGACTGCTA 54121

Qy   1812 tttcaccttcacgaaacacacaaaaacacagcgaattgtagcacctcaaagattcgatag 1871
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54122 TTTCACCTTCACGAAACACACAAAAACACAGCGAATTGTAGCACCTCAAAGATTCGATAG 54181

Qy   1872 ctttttgtcatagtccttagtcttaactcgtatttatttttcgtacggttgtcgagctcaa 1931
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54182 CTTTTTGTCATAGTCCTTAGTCTTAACTCGTATTTATTTTCGTACGGTTGTCGAGCTCAA 54241

Qy   1932 aaataaaatcaaattaagctaaaa 1955
          ||||||||||||||||||||||| |
Db  54242 AAATAAAATCAAATTAAGCTAAGA 54265
```

```
RESULT   3
AE003824
LOCUS       AE003824   249001 bp    DNA             INV       22-MAR-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386047 section 34
            of 52, complete sequence.
ACCESSION   AE003824 AE002787
VERSION     AE003824.1  GI:7303542
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 249001)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
```

```
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 249001)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
FEATURES             Location/Qualifiers
     source          1. .249001
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
     mRNA            complement(join(<5140. .5357,5490. .6855,6912. .7040,
                     7095. .>7172))
                     /gene="CG8298"
                     /product="CT24549"
                     /db_xref="FLYBASE:FBan0008298"
                     /db_xref="FLYBASE:FBgn0033673"
                     /evidence=not_experimental
     gene            complement(<5140. .>7172)
                     /gene="CG8298"
                     /db_xref="FLYBASE:FBan0008298"
                     /db_xref="FLYBASE:FBgn0033673"
                     /evidence=not_experimental
     CDS             complement(join(5140. .5357,5490. .6855,6912. .7040,
                     7095. .7172))
                     /gene="CG8298"
                     /note="CG8298 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0008298"
                     /db_xref="FLYBASE:FBgn0033673"
                     /evidence=not_experimental
                     /protein_id="AAF58597.1"
                     /db_xref="GI:7303543"
                     /translation="MSNGRYGRFGALIGVACVSSTHCRFLIYSTKNAEVLAYHELKLR
                     QIVHQAGWMEYDPSEIWKNTQECIETAYKNLVILEINPRDIIAIGIVNQRGTSVLWNL
                     ETGQPLHNAIGWSDCRSTPILKTLLHNVRHNVDYVRYRSGLPLSSCFSALKIRWLMDH
                     VPAVATAIEENKCLFGTLDSWLLWNLTGGVEMGVHSTDITNAHYTSLMNVSTEQWDPK
                     LCQFFRLPLNILPRIRSNSEIFGYVLEGPLHGTPIAAMMGEQPASLLGQLCVKAGQNV
                     CTLDDSCFVLLNTGREKLDSANGLITGIAHKLGEKAATNYTLEGAISNAGSTVTWLRD
                     KLQINTEINSNDNVVESLNTFIGENSMISSSCSSSMLNAECGLAAKRSEITFVPAFHG
                     MYAPYWRHDARGIILGLTSQTTAENITQAAYEATGFQIFEVLQAFKRDTPNWDRSSMQ
                     PVLTFGGDYAENLHLVQFIADIIGYMLERPQTTSPAGLGVMITAGVTMKVVSLEHAVK
                     MYTPPTDVFSPTTTKNRRELLYKRWAYAVKKCLHWNNYETYEADIELFAQRELDPNLP
                     VRRSVPGSIFLTTSFAMIILANYLKSNPLK"
     mRNA            join(7688. .7799,7859. .8060,8126. .8420,8921. .9141,
                     9196. .9395,9460. .>9550)
                     /gene="CG8972"
                     /product="CT25796"
                     /db_xref="FLYBASE:FBan0008972"
                     /db_xref="FLYBASE:FBgn0033672"
     gene            <7688. .>9550
                     /gene="CG8972"
                     /db_xref="FLYBASE:FBan0008972"
                     /db_xref="FLYBASE:FBgn0033672"
     CDS             join(7753. .7799,7859. .8060,8126. .8420,8921. .9141,
                     9196. .9395,9460. .9550)
                     /gene="CG8972"
                     /note="CG8972 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0008972"
                     /db_xref="FLYBASE:FBgn0033672"
                     /protein_id="AAF58598.1"
                     /db_xref="GI:7303544"
                     /translation="MLMSRALCRSWLPQVARRCHANVNVPILRINSGHPAARSCRQIH
                     SNRKQSSNLKPTTGEPAAAEQNTPVPVNNVIKAVAFTGAFTVGCFAGATILEYENTRS
                     LILEKARQARFGWWQSRSLADRDYWTQIKQDIRRHWDSLTPGDKMFAPILLCNLVAFA
                     MWRVPALKSTMITYFTSNPAAKVVCWPMFLSTFSHYSAMHLFANMYVMHSFANAAAVS
                     LGKEQFLAVYLSAGVFSSLMSVLYKAATSQAGMSLGASGAIMTLLAYVCTQYPDTQLS
                     ILFLPALTFSAGAGIKVLMGIDFAGVVMGWKFFDHAAHLGGAMFGIFWATYGAQIWAK
                     RIGLLNYYHDLRRTKQK"
     mRNA            join(11156. .11206,11355. .11603,12112. .>12993)
                     /gene="RnrS"
                     /product="CT25800"
                     /db_xref="FLYBASE:FBan0008975"
                     /db_xref="FLYBASE:FBgn0011704"
                     /evidence=not_experimental
     gene            <11156. .>12993
                     /gene="RnrS"
                     /note="CG8975"
                     /map="48D-48D"
                     /db_xref="FLYBASE:FBan0008975"
                     /db_xref="FLYBASE:FBgn0011704"
                     /evidence=not_experimental
     CDS             join(11186. .11206,11355. .11603,12112. .12993)
                     /gene="RnrS"
                     /note="RnrS gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0008975"
                     /db_xref="FLYBASE:FBgn0011704"
                     /evidence=not_experimental
                     /protein_id="AAF58599.1"
                     /db_xref="GI:7303545"
                     /translation="MEKFSLKSPSKKILTDSTNNVRKMSIGHEANGQLAKESSTVNGI
                     GKSANSLMEKSVTPFDPSLEPLLRENPRRFVIFPIQYHDIWQMYKKAEASFWTVEEVD
                     LSKDLTDWHRLKDDERHFISHVLAFFAASDGIVNENLVERFSQEVQITEARCFYGFQI
                     AMENVHSEMYSVLIDTYIRDPHQREYLFNAIETMPAVKRKADWALSWISSKSANFGER
                     IIAFAAVEGIFFSGSFASIFWLKKRGLMPGLTFSNELISRDEGLHCDFAVLMFQHLVQ
                     RPKRERIIEIIRDAVAIEQEFLTDALPVNLIGMNCDLMSQYIEFVADRLLVELGVGKI
                     YNTKNPFNFMEMISLDGKTNFFEKKVGEYQRMGVVSNPLDNVFTLDADF"
     mRNA            join(<13698. .13986,14051. .15157,15956. .16090,16149. .17250,
                     17502. .17536,17592. .>18694)
                     /gene="CG8976"
                     /product="CT25806"
                     /db_xref="FLYBASE:FBan0008976"
                     /db_xref="FLYBASE:FBgn0033671"
                     /evidence=not_experimental
     gene            <13698. .>18694
                     /gene="CG8976"
                     /db_xref="FLYBASE:FBan0008976"
                     /db_xref="FLYBASE:FBgn0033671"
                     /evidence=not_experimental
     CDS             join(13698. .13986,14051. .15157,15956. .16090,16149. .17250,
                     17502. .17536,17592. .18694)
                     /gene="CG8976"
                     /note="CG8976 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0008976"
                     /db_xref="FLYBASE:FBgn0033671"
                     /evidence=not_experimental
                     /protein_id="AAF58600.1"
```

```
                /db_xref="GI:7303546"
                /translation="MDRRWNPEKIEEQSSLLAAEYSSSSGAAASGSEDETVTTSGALR
                SKAKLRKRQLRNRMPLPRMLRRLGCYTLSAFLICGLLLVYLPLVYLDVHKRSAGLPDW
                TSETSRSIADYLDIGLSSGVIVPKDFCRNKTFLVIAVCTGVDNFIQRHTIRETWGNTT
                EFNYPAFGKLHGHLKGHYLPPLPDRLKMYGDYLSGEGQSLTASVRIVFIVGRQKDEAM
                LGNETLNRIHIESEKYNDIIQENFVDSYNNLTLKSVMALKHISRSCFNTAVYFLKCDD
                DTFVNIPNLLNFLLGGTIPLYNDTLDYHDRSTYLVTAPQTRLKASSDVLYGHQFCNVV
                PVSEVSSKWYMPSYMYKPESYPKYLSGAGYLMSIDVVQRLFEASLNTTLVYLEDVYIT
                GLCAQKAKINRHHHPLFSFAHSKQMCAFKGTITQHQLKDDSMVSAWNYVSNYSIKCPP
                PGRYFSQVRLRKRPIYELKRLYRIVARIWLFIRRYKAIIILAVFLAYRIFNFKPDPSK
                RTASLAGWSRDAPRSVADYLDPTKDTALIVPRHFCRSKALLTIAVCSYVHHFERRRAI
                RKLWGNFTDFNYSVFVKLHGHLKGRYQDVLPERLKLYSEYLSGEGDSLRASIRLVFIV
                GRRNLASLLENEAVAIEAQKYNDVIQENFIDTYNNLTIKAVMALKHITQSCLNTTAFY
                FKCDDDTFVNVPNILHFLLGGTIPVNVVTAGFHYGNTYEVTSPRKRLTARREMMYGRQ
                HCNVPPATNKLNKWYMPSYMFRGGVYPRYLCGSGYLLSIDVVPRLYKASLGTRIVHLE
                DMFVTGLCAEKAGIKRTNHPLFRSSYPYEGDEQCALKGSFTVHRAKDNVMWEAWYRVT
                NFSSKCPPPQKDFHLRLPKMQSPRPVPDVHDRKAELSGWEREKPLKIGDYLDQGKNTA
                LIVPRDFCKSKALLVIAVCSGLDHFEQRSAIRQTWGNTKSFNYGEFVRLHGHLEGKYL
                SVMPGRLKLYSMYLSGLDDSLTAKIRIVFILGRSKNDSKRELDKLFRESIQHNDIIQE
                NFVDSYHNLTLKSVMALKHISQSCADRAAFFLKCDDDTFVNVPNLLHFLLGGTIPLYK
                DTVGYHSRTTYKVKSPWNRLNGSRGLMYGHKFCNMKTVDDVKSPWYMPYYMFKGAKYP
                KYLSGTGYLMSIDVVKRLYAEALTTSLVHLEDVFVTGICAKKAGIRRRHQPLFNYVHG
                KPLCIFKGTITMHPVPLHSMLDAWAFVSNYSIKCLPPRKHFVKTELSKKSHC"
     mRNA       complement(join(19346. .19752,19879. .20106,20173. .21066,
                21358. .21831,21914. .22864,23382. .24697,24816. .25284,
                25347. .>25744))
                /gene="CG8290"
                /product="CT24533"

  Query Match             91.6%;  Score 1806.4;  DB 58;  Length 249001;
  Best Local Similarity   94.2%;  Pred. No. 0;
  Matches 1945;  Conservative    0;  Mismatches    1;  Indels  118;  Gaps    2;

Qy      10 gattaatgctgatttctggtctggactacacagcattgctggtataaggagtcgggacca 69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152260 GATTAATGCTGATTTCTGGTCTGGACTACACAGCATTGCTGGTATAAGGAGTCGGGACCA 152319

Qy      70 gaggagtaagatttcgggaaggaatcccgtccggtagggactactagcattcgcaagtga 129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152320 GAGGAGTAAGATTTCGGGAAGGAATCCCGTCCGGTAGGGACTACTAGCATTCGCAAGTGA 152379

Qy     130 cgtccagcaaccggaggacccccaactgtagaatccgcatcaccatcctaatcccaacaa 189
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152380 CGTCCAGCAACCGGAGGACCCCCAACTGTAGAATCCGCATCACCATCCTAATCCCAACAA 152439

Qy     190 accaatgacatcttgagacctcaccagccatggatcccttcgtgttcttcatagtactgg 249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152440 ACCAATGACATCTTGAGACCTCACCAGCCATGGATCCCTTCGTGTTCTTCATAGTACTGG 152499

Qy     250 catcgctttatggcgttctttacttttcgaccgcttcttca------------------ 291
           |||||||||||||||||||||||||||||||||||||||
Db 152500 CATCGCTTTATGGCGTTCTTTACTTTTTCGACCGCTTCTTCAAGGTGTAGTATATCCAGC 152559

Qy     292 ---------------------------------agagttgcatgcactacccgtacgatg 318
                                           ||||||||||||||||||||||||||
Db 152560 CAAAGTACGTCCAGATACTTAATGTAATCCCTTAGAGTTGCATGCACTACCCGTACGATG 152619

Qy     319 ccttcctcaagaacaccgggctgagtataaatttcatgagcctccactggcacacgagtg 378
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152620 CCTTCCTCAAGAACACCGGGCTGAGTATAAATTTCATGAGCCTCCACTGGCACACGAGTG 152679

Qy     379 cctttaacaggaccctcctacgctggggatctgccggtaacagctgcacccggagagtaa 438
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152680 CCTTTAACAGGACCCTCCTACGCTGGGGATCTGCCGGTAACAGCTGCACCCGGAGAGTAA 152739

Qy     439 tgatcaccagctttaatgtaggagtcctggtcacctttttctctgctcccgatcggtctga 498
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152740 TGATCACCAGCTTTAATGTAGGAGTCCTGGTCACCTTTTCTCTGCTCCCGATCGGTCTGA 152799

Qy     499 tcctgctcattgccactatcttcagcagtggtgaacaagacagctcttcgtctgtatcct 558
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152800 TCCTGCTCATTGCCACTATCTTCAGCAGTGGTGAACAAGACAGCTCTTCGTCTGTATCCT 152859

Qy     559 cgcccgttggagtccctgtgcagctggaaattctactgcccggcgtcaacttgccgttgg 618
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152860 CGCCCGTTGGAGTCCCTGTGCAGCTGGAAATTCTACTGCCCGGCGTCAACTTGCCGTTGG 152919

Qy     619 aggagatcggatactacatcacaacccttgtgctctgcttggtggtgcacgagatgggac 678
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152920 AGGAGATCGGATACTACATCACAACCCTTGTGCTCTGCTTGGTGGTGCACGAGATGGGAC 152979

Qy     679 acgccctggccgctgtgatggaggatgtgcctgtcaccgggtttggaataa--------- 729
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152980 ACGCCCTGGCCGCTGTGATGGAGGATGTGCCTGTCACCGGGTTTGGAATAAAGGTTAGAA 153039

Qy     730 ------------------------------------------------------------ag 731
                                                                        ||
Db 153040 TGATATAGCGTACTTTACAACCTAATACGCTTACTAATTAGGTAAATATTAATATTTCAG 153099

Qy     732 ttcatcttctgcctgccgttagcatacacggagctctcccacgaccacttaaacagtcta 791
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 153100 TTCATCTTCTGCCTGCCGTTAGCATACACGGAGCTCTCCCACGACCACTTAAACAGTCTA 153159

Qy     792 cgttggttccgcaagctacgtgttctgtgcgctggaatctggcataattttgtgttcgct 851
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 153160 CGTTGGTTCCGCAAGCTACGTGTTCTGTGCGCTGGAATCTGGCATAATTTTGTGTTCGCT 153219

Qy     852 ggcgtgtgctatctcttaatctcaacggtgggaatcactatgtcacctttgtacgcttac 911
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 153220 GGCGTGTGCTATCTCTTAATCTCAACGGTGGGAATCACTATGTCACCTTTGTACGCTTAC 153279

Qy     912 aaccaacacgtagtggtcactgaactaacaaggaaatccccgctgaggggagagcgcggc 971
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 153280 AACCAACACGTAGTGGTCACTGAACTAACAAGGAAATCCCCGCTGAGGGGAGAGCGCGGC 153339

Qy     972 ttgcaagtggacaatcaaataacccaagtaaacggctgcccagtaaacagcgaggagagt 1031
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 153340 TTGCAAGTGGACAATCAAATAACCCAAGTAAACGGCTGCCCAGTAAACAGCGAGGAGAGT 153399

Qy    1032 tgggtgacatgcctgcagaactctctgaagctcaagccgggctactgtgtgagtgcggac 1091
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 153400 TGGGTGACATGCCTGCAGAACTCTCTGAAGCTCAAGCCGGGCTACTGTGTGAGTGCGGAC 153459

Qy    1092 ttcgtgcagcttaacgacgaaagcagcgccatctcacatcatagcattgatggtcagcta 1151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 153460 TTCGTGCAGCTTAACGACGAAAGCAGCGCCATCTCACATCATAGCATTGATGGTCAGCTA 153519

Qy    1152 cagtgctgtgatgaactaaatccgaacgtaagctgcttcgaggtggtggaggacgcaaat 1211
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 153520 CAGTGCTGTGATGAACTAAATCCGAACGTAAGCTGCTTCGAGGTGGTGGAGGACGCAAAT 153579

Qy    1212 ggagatgtgccggtggagctgccgcagcatgtatgtctcaatgtgcgccgcactttggag 1271
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 153580 GGAGATGTGCCGGTGGAGCTGCCGCAGCATGTATGTCTCAATGTGCGCCGCACTTTGGAG 153639

Qy    1272 gaggtctccgagcactgctcgtccggagtttgcaacgagggattctgcctacgaccgctt 1331
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 153640 GAGGTCTCCGAGCACTGCTCGTCCGGAGTTTGCAACGAGGGATTCTGCCTACGACCGCTT 153699

Qy    1332 atacgaaatatcactgccataatgacgttcaagcgacagaattttcgcggagagaagctg 1391
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 153700 ATACGAAATATCACTGCCATAATGACGTTCAAGCGACAGAATTTTCGCGGAGAGAAGCTG 153759

Qy    1392 ccgccggtgatctatgtgggccatccatgggatgtcactcgaactgtggaggtatccgcc 1451
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 153760 CCGCCGGTGATCTATGTGGGCCATCCATGGGATGTCACTCGAACTGTGGAGGTATCCGCC 153819

Qy    1452 tttgtgccgagatatagcttattaaaggcagcctggccggatgcctggctgctgctcctc 1511
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 153820 TTTGTGCCGAGATATAGCTTATTAAAGGCAGCCTGGCCGGATGCCTGGCTGCTGCTCCTC 153879

Qy    1512 aagtataacgtggtcttcagcataggattggcgttgatcaatgccattccctgctttggt 1571
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 153880 AAGTATAACGTGGTCTTCAGCATAGGATTGGCGTTGATCAATGCCATTCCCTGCTTTGGT 153939

Qy    1572 ttcgatggcgcccacattaccagcaccgtgatacacagcttcttggtgggcagagtggat 1631
```

```
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 153940 TTCGATGGCGCCCACATTACCAGCACCGTGATACACAGCTTCTTGGTGGGCAGAGTGGAT 153999

Qy    1632 cagcatgccaagagagatatcatctcgttgataatcaccagcgtgggttcccttctcttt 1691
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 154000 CAGCATGCCAAGAGAGATATCATCTCGTTGATAATCACCAGCGTGGGTTCCCTTCTCTTT 154059

Qy    1692 gcactggccctgcttaaggtggcctggttgagttttctgcgacccctgctttaagaactg 1751
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 154060 GCACTGGCCCTGCTTAAGGTGGCCTGGTTGAGTTTTCTGCGACCCCTGCTTTAAGAACTG 154119

Qy    1752 aaatggaaaactgaaatggatcctgggagttcaactccctgcaaagacgctagactgcta 1811
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 154120 AAATGGAAAACTGAAATGGATCCTGGGAGTTCAACTCCCTGCAAAGACGCTAGACTGCTA 154179

Qy    1812 tttcaccttcacgaaacacacaaaaacacagcgaattgtagcacctcaaagattcgatag 1871
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 154180 TTTCACCTTCACGAAACACACAAAAACACAGCGAATTGTAGCACCTCAAAGATTCGATAG 154239

Qy    1872 ctttttgtcatagtccttagtcttaactcgtatttatttcgtacggttgtcgagctcaa 1931
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 154240 CTTTTTGTCATAGTCCTTAGTCTTAACTCGTATTTATTTCGTACGGTTGTCGAGCTCAA 154299

Qy    1932 aaataaaatcaaattaagctaaaa 1955
             |||||||||||||||||||||| |
Db 154300 AAATAAAATCAAATTAAGCTAAGA 154323


RESULT   4
AF019611
LOCUS       AF019611     1732 bp    mRNA            ROD       09-JAN-1998
DEFINITION  Cricetulus griseus S2P mRNA, complete cds.
ACCESSION   AF019611
VERSION     AF019611.1  GI:2745730
KEYWORDS    .
SOURCE      Chinese hamster.
  ORGANISM  Cricetulus griseus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
REFERENCE   1  (bases 1 to 1732)
  AUTHORS   Rawson,R.B., Zelenski,N.G., Nijhawan,D., Ye,J., Sakai,J.,
            Hasan,M.T., Chang,T.Y., Brown,M.S. and Goldstein,J.L.
  TITLE     Complementation cloning of S2P, a gene encoding a putative
            metalloprotease required for intramembrane cleavage of SREBPs
  JOURNAL   Mol. Cell 1 (1), 47-57 (1997)
  MEDLINE   98324087
REFERENCE   2  (bases 1 to 1732)
  AUTHORS   Rawson,R.B., Zelenski,N.G., Nijhawan,D., Ye,J., Sakai,J.,
            Hasan,M.T., Chang,T.Y., Brown,M.S. and Goldstein,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-1997) Molecular Genetics, UT Southwestern Medical
            Center, 5323 Harry Hines Blvd, Dallas, TX 75235, USA
FEATURES             Location/Qualifiers
     source          1..1732
                     /organism="Cricetulus griseus"
                     /db_xref="taxon:10029"
                     /chromosome="X"
                     /cell_line="CHO-7"
                     /note="cell line is a subline of CHO K1 that has been
                     selected for growth in lipid-deficient medium"
     CDS             100..1632
                     /codon_start=1
                     /product="S2P"
                     /protein_id="AAC53526.1"
                     /db_xref="GI:2745731"
                     /translation="MIPVSLVVVVVGGWTAVYLADLVLKSSVYFKHSYEDWLEKNGLS
                     ISPFHIRWQTSVFNRAFYSWGRRKARMLYQWFNFGMVFGVIAMFSSFFLLGKTLMQTL
                     AQMMADSPSSSSSSSSSSSSSSSSIHNEQVLQVVVPGINLPVNQLTYFFAAVLISGV
                     VHEIGHGIAAIREQVRFNGFGIFLFIIYPGAFVDLFTTHLQLISPVQQLRIFCAGIWH
                     NFVLALLGILALVLLPVILLPFYYTGVGVLITEVAEDSPAIGPRGLFVGDLVTHLQDC
                     PVTNVQDWNECLDTIAYEPQIGYCISASTLQQLSFPVRAYKRLDGSTECCNNHSLTDV
                     CFSYRNNFNKRLHTCLPARKAVEATQVCRTNKDCKTSSSSSFCIVPSLETHTRLIKVK
                     HPPQIDMLYVGHPLHLHYTVSITSFIPRFNFLSIDLPVIVETFVKYLISLSGALAIVN
                     AVPCFALDGQWILNSFLDATLTSVIGDNDVKDLIGFFILLGGSVLLAANVTLGLWMVT
                     AR"
BASE COUNT      414 a    375 c    381 g    562 t
ORIGIN

  Query Match             4.5%;  Score 88.6;  DB 12;  Length 1732;
  Best Local Similarity   48.3%;  Pred. No. 4.4e-14;
  Matches  247;  Conservative    0;  Mismatches  264;  Indels    0;  Gaps    0;

Qy       542 ctcttcgtctgtatcctcgcccgttggagtccctgtgcagctggaaattctactgcccgg 601
              ||| || ||     |||||  |  ||         |  | || || ||  |  | || ||
Db       459 CTCCTCTTCCTCTTCCTCCTCGATTCACAATGAACAGGTGTTGCAAGTTGTGGTTCCTGG 518

Qy       602 cgtcaacttgccgttggaggagatcggatactacatcacaacccttgtgctctgcttggt 661
              | || || ||  |  |  || |   |||| | || |  |  |  |   |  |    ||
Db       519 TATAAATTTACCTGTCAATCAGCTGACTTACTTCTTCGCTGCAGTCCTCATTAGTGGTGT 578

Qy       662 ggtgcacgagatgggacacgccctggccgctgtgatggaggatgtgcctgtcaccgggtt 721
              || || || || ||||||   | || ||| | | |||  | || |   |  | || ||
Db       579 TGTACATGAAATTGGACACGGTATAGCAGCTATTAGGGAACAAGTTCGATTTAATGGCTT 638

Qy       722 tggaataaagttcatcttctgcctgccgttagcatacacggagctctcccacgaccactt 781
             |||||      || || |       ||     ||  |      || || |  |   || ||
Db       639 TGGAATTTTTCTCTTCATTATTTATCCTGGCGCTTTTGTTGATCTGTTCACCACACATTT 698

Qy       782 aaacagtctacgttggttccgcaagctacgtgttctgtgcgctggaatctggcataattt 841
               |   | ||      |||    ||||| |  |  | || || || || |||||||||
Db       699 GCAACTTATATCACCAGTCCAGCAGCTAAGGATATTTTGTGCAGGTATATGGCATAATTT 758

Qy       842 tgtgttcgctggcgtgtgctatctcttaatctcaacggtgggaatcactatgtcacttt 901
             |||  | |||  ||  |        | |        |   | | | | | |   || ||
Db       759 TGTCCTTGCTCTCCTTGGAATTTTAGCTCTTGTTCTCCTCCCAGTAATTCTCTTGCCATT 818

Qy       902 gtacgcttacaaccaacacgtagtggtcactgaactaacaaggaaatccccgctgagggg 961
              |||             |||| |  |||| |||  |  ||  | | | || ||    |  ||
Db       819 TTACTACACTGGAGTTGGGGTACTTATCACGGAAGTCGCAGAGGACTCACCTGCCATTGG 878

Qy       962 agagcgcggcttgcaagtggacaatcaaataacccaagtaaacggctgcccagtaaacag 1021
              |    | ||| |    ||||   | |    | || ||  |  | |  ||||| || |  |
Db       879 ACCCAGAGGCCTTTTTGTGGGAGACCTTGTCACTCATCTTCAGGATTGCCCTGTTACTAA 938

Qy      1022 cgaggagagttgggtgacatgcctgcagaac 1052
              || |  |     ||||    |||  |  | |
Db       939 CGTGCAAGATTGGAATGAATGTTTAGATACC 969


RESULT   5
AF019612
LOCUS       AF019612     1759 bp    mRNA            PRI       09-JAN-1998
DEFINITION  Homo sapiens S2P mRNA, complete cds.
ACCESSION   AF019612
VERSION     AF019612.1  GI:2745732
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1759)
  AUTHORS   Rawson,R.B., Zelenski,N.G., Nijhawan,D., Ye,J., Sakai,J.,
            Hasan,M.T., Chang,T.Y., Brown,M.S. and Goldstein,J.L.
  TITLE     Complementation cloning of S2P, a gene encoding a putative
            metalloprotease required for intramembrane cleavage of SREBPs
  JOURNAL   Mol. Cell 1 (1), 47-57 (1997)
  MEDLINE   98324087
REFERENCE   2  (bases 1 to 1759)
  AUTHORS   Rawson,R.B., Zelenski,N.G., Nijhawan,D., Ye,J., Sakai,J.,
            Hasan,M.T., Chang,T.Y., Brown,M.S. and Goldstein,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-1997) Molecular Genetics, UT Southwestern Medical
            Center, 5323 Harry Hines Blvd, Dallas, TX 75235, USA
FEATURES             Location/Qualifiers
     source          1..1759
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /cell_line="HeLa"
     CDS             100. .1659
                     /codon_start=1
                     /product="S2P"
                     /protein_id="AAC51937.1"
                     /db_xref="GI:2745733"
                     /translation="MIPVSLVVVVVGGWTVVYLTDLVLKSSVYFKHSYEDWLENNGLS
                     ISPFHIRWQTAVFNRAFYSWGRRKARMLYQWFNFGMVFGVIAMFSSFFLLGKTLMQTL
                     AQMMADSPSSYSSSSSSSSSSSSSSSSSSSSSLHNEQVLQVVVPGINLPVNQLTYF
                     FTAVLISGVVHEIGHGIAAIREQVRFNGFGIFLFIIYPGAFVDLFTTHLQLISPVQQL
                     RIFCAGIWHNFVLALLGILALVLLPVILLPFYYTGVGVLITEVAEDSPAIGPRGLFVG
                     DLVTHLQDCPVTNVQDWNECLDTIAYEPQIGYCISASTLQQLSFPVRAYKRLDGSTEC
                     CNNHSLTDVCFSYRNNFNKRLHTCLPARKAVEATQVCRTNKDCKKSSSSSFCIIPSLE
                     THTRLIKVKHPPQIDMLYVGHPLHLHYTVSITSFIPRFNFLSIDLPVVVETFVKYLIS
                     LSGALAIVNAVPCFALDGQWILNSFLDATLTSVIGDNDVKDLIGFFILLGGSVLLAAN
                     VTLGLWMVTAR"
BASE COUNT      410 a    399 c    380 g    570 t
ORIGIN


  Query Match             4.4%;  Score 86;  DB 11;  Length 1759;
  Best Local Similarity   48.0%;  Pred. No. 2.5e-13;
  Matches  245;  Conservative    0;  Mismatches  265;  Indels    0;  Gaps    0;

Qy    543 tcttcgtctgtatcctcgcccgttggagtccctgtgcagctggaaattctactgcccggc 602
          || || |||   || ||  |  ||           |  | |  || || |  | || ||
Db    487 TCATCTTCTTCCTCTTCCTCGCTTCACAATGAACAGGTGTTACAAGTTGTGGTTCCTGGT 546

Qy    603 gtcaacttgccgttggaggagatcggatactacatcacaacccttgtgctctgcttggtg 662
          | || || ||  |  |  |  |    || | | ||||  |  || |  |  |    ||
Db    547 ATAAATTTACCCGTCAATCAACTGACCTATTTCTTCACGGCAGTTCTCATTAGTGGTGTT 606

Qy    663 gtgcacgagatgggacacgccctggccgctgtgatggaggatgtgcctgtcaccgggttt 722
          || || || || ||||| |   | || ||| | | |||  | || |   | |  || |||
Db    607 GTACATGAAATTGGACATGGGATAGCAGCTATTAGGGAACAAGTTCGATTTAATGGCTTT 666

Qy    723 ggaataaagttcatcttctgcctgccgttagcatacacggagctctcccacgaccactta 782
          || ||     || || |       ||   |||||     || || | |  |    || ||
Db    667 GGGATTTTTCTCTTCATTATTTATCCTGGAGCATTTGTTGATCTGTTCACCACTCATTTG 726

Qy    783 aacagtctacgttggttccgcaagctacgtgttctgtgcgctggaatctggcataatttt 842
          |   | ||       |||   ||||| |  |  | || || || ||||||||||||||||
Db    727 CAACTTATATCGCCAGTCCAGCAGCTAAGGATATTTTGTGCAGGTATCTGGCATAATTTT 786

Qy    843 gtgttcgctggcgtgtgctatctcttaatctcaacggtgggaatcactatgtcacctttg 902
          ||  | ||   | || |   | |       |       | | | | | |  || ||
Db    787 GTCCTTGCACTCTTGGGTATTTTAGCTCTTGTTCTCCTCCCAGTAATTCTCTTGCCATTT 846

Qy    903 tacgcttacaaccaacacgtagtggtcactgaactaacaaggaaatccccgctgagggga 962
          |||            ||  |  |||||||| |  |   | | || ||    |  |||
Db    847 TACTACACTGGAGTTGGGGTGCTCATCACTGAAGTTGCTGAGGACTCTCCTGCCATTGGA 906

Qy    963 gagcgcggcttgcaagtggacaatcaaataacccaagtaaacggctgcccagtaaacagc 1022
            | ||| |    ||||   | |   | |||||  || | |  || || || |  |
Db    907 CCCAGAGGCCTTTTTGTGGGAGACCTTGTCACCCATCTACAGGATTGTCCTGTTACTAAT 966

Qy   1023 gaggagagttgggtgacatgcctgcagaac 1052
          | | |  ||||     |||  |  | | |
Db    967 GTGCAAGATTGGAATGAATGTTTAGATACC 996


RESULT   6
DDU25143/c
LOCUS       DDU25143     3946 bp    DNA             INV       29-JUL-1995
DEFINITION  Dictyostelium discoideum cAMP-inducible prespore protein gene,
            complete cds.
ACCESSION   U25143
VERSION     U25143.1  GI:915200
KEYWORDS    .
SOURCE      slime mold.
```

```
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 3946)
  AUTHORS   Agarwal,A., Sloger,M.S., Oyama,M. and Blumberg,D.D.
  TITLE     Analysis of a novel cyclic Amp inducible prespore gene in
            Dictyostelium discoideum: evidence for different patterns of cAMP
            regulation
  JOURNAL   Differentiation 57 (3), 151-162 (1994)
  MEDLINE   95080502
REFERENCE   2  (bases 1 to 3946)
  AUTHORS   Blumberg D.,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-APR-1995) Daphne D. Blumberg, University of Maryland
            Baltimore County, Biological Sciences, Baltimore, MD 21228, USA
FEATURES             Location/Qualifiers
     source          1. .3946
                     /organism="Dictyostelium discoideum"
                     /strain="Ax3"
                     /db_xref="taxon:44689"
                     /clone="D7"
     exon            1008. .1226
                     /number=1
     CDS             join(1161. .1226,1412. .3898)
                     /codon_start=1
                     /product="cAMP-inducible prespore protein"
                     /protein_id="AAA73514.1"
                     /db_xref="GI:915201"
                     /translation="MYSKKYTSFVIVLILSCIISTCTSNQISEDVGKAINQKLNENIE
                     KVEDVVVQFENVANQVIEELKVEHQRQELLGEVKHDSLRDSATNYIWGLLDKIQSYLP
                     KDNNKVSKVEEAFSSGQNNNIGSSIGDSTGASTSPQFQSINGLSGASQSSGSSTGGTG
                     DSDSKTTNEAIIFSSKVSTTDRQESIIGQVAITAKDSLGATITGLGGVSSTAKVGGQI
                     TNGRAQGQVITGGDNTGTVGRGAVTTASAVANTVGEFLGGSRTGGSSSAGTVGNVISD
                     SYTSIGKIASGNGNSLSETIGTTGDTLAHTFAGTDSVGVTGFHIITKTFNLIAGGKFN
                     SDQQYIDKSYGSIPSQDNEEIKKRLQSAHQQLQEQSPAIYQSMKSEDLKNLDDEVIRN
                     TLKEMQIQRENQDIGQQNQEDKEQLIDLQNREPGLYKNQQDLKQEKRANQQELINYEL
                     NLQEDQEQYELLLDQLYDEQQQQPQKVSNKQQLQEQQINSPEDIQYQLNHLNQPFQDD
                     YHNDQTEELKDDDYNFNDQQINNGQFENNVEEFPDLNDANDNFEQVNNNNNNINNNNN
                     NNFKVDKSKKSAQQVEIALENERLYLQEVEDAPERLYEEIHNSNLNKAVQEAEEIERQ
                     QNGNGSPAVNSHKIVTQMGNVENNEITNEQDAIEQAQHLQRIENGEDIDEYEAAQHEV
                     AFLDENEQDAAKLQNEMDQLELLLEENEEQFSNKNLQQLNENENQQQQQGQQQVQAT
                     NVEQQIVEQLKVIKEFQRQDQQKKQKIQQENDAVYLSDVEKAQLELDAELAKNNKQEN
                     QDENLVQEKQQSPDQIKNQLKNIQHEQQIQEQQDKLNQEKNQQLLEQEQNQK"
     intron          1227. .1411
                     /number=1
     exon            1412. .>3898
                     /number=2
     polyA_signal    3894. .3898
BASE COUNT     1631 a    478 c    502 g   1335 t
ORIGIN


  Query Match             2.4%;  Score 46.4;  DB 33;  Length 3946;
  Best Local Similarity   63.4%;  Pred. No. 0.06;
  Matches   71;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy   1859 aaagattcgatagctttttgtcatagtccttagtcttaactcgtatttatttttcgtacgg 1918
           |||||    |||| | |||| |      ||| | || | |  | ||| ||||  |
Db   1047 AAAGAAAAAAATAGTTATTTGATAGTAAATTTATTATTTAATTTTTTTTTTTTTTTCTTT 988

Qy   1919 ttgtcgagctcaaaaataaaatcaaattaagctaaaaaaaaaaaaaaaaaaa 1970
          || |  |  | | ||| ||||  |||||||  |||||||||||||||||| ||
Db    987 TTTTTTAAATAATAAAAAAAAAAAAATTAAATTAAAAAAAAAAAAAAAAGAA 936


RESULT   7
I66494
LOCUS       I66494       7218 bp    DNA             PAT       28-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
```

```
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN


  Query Match             2.3%;  Score 46;  DB 5;  Length 7218;
  Best Local Similarity   4.7%;  Pred. No. 0.088;
  Matches   19;  Conservative  215;  Mismatches  170;  Indels    0;  Gaps    0;

Qy     153 aactgtagaatccgcatcaccatcctaatcccaacaaaccaatgacatcttgagacctca 212
           | ||   || ::: : :: :: :::: :::: :  ::  :  : ::::    ::::
Db    1057 AGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1116

Qy     213 ccagccatggatcccttcgtgttcttcatagtactggcatcgctttatggcgttctttac 272
           ::  :: :   ::::::: : :::::: :  : ::  : :: :::: :  : :::::: :
Db    1117 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1176

Qy     273 tttttcgaccgcttcttcaagagttgcatgcactacccgtacgatgccttcctcaagaac 332
           ::::::  :: :::::::     :: : : : :: ::: : :  : :::::::     :
Db    1177 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1236

Qy     333 accgggctgagtataaatttcatgagcctccactggcacacgagtgcctttaacaggacc 392
           ::   ::   : :   :::: :  ::::: ::  : : :   : :::::  :   ::
Db    1237 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1296

Qy     393 ctcctacgctggggatctgccggtaacagctgcacccggagagtaatgatcaccagcttt 452
           ::::: : ::    ::: ::  :  :  :: : :::     :  :  :: ::  ::::
Db    1297 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1356

Qy     453 aatgtaggagtcctggtcacctttctctgctcccgatcggtctgatcctgctcattgcc 512
           : :    ::::  :: :::::::::: ::::  ::  :::  :::: ::: :: ::
Db    1357 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1416

Qy     513 actatcttcagcagtggtgaacaagacagctcttcgtctgtatc 556
           :: :::::  :  :  :| || | |    |||   ||| || |
Db    1417 YYYYYYYYYYYYYYYYYGTACCAAATTCTTCTATCTCTTTAAC 1460


RESULT   8
PFMAL3P8
LOCUS       PFMAL3P8   108908 bp    DNA             INV       17-MAR-2000
DEFINITION  Plasmodium falciparum MAL3P8, complete sequence.
ACCESSION   AL034560 AL008968 AL010134 AL010151 AL010159 AL010171 AL010188
            AL010209 AL021886 AL021887 Z97346
VERSION     AL034560.3  GI:4725993
KEYWORDS    HTG; alpha/beta hydrolase; guanine nucleotide-binding protein;
            long-chain-fatty-acid-CoA ligase; R-FA3; rep11; rep20; rifin;
            Serine/threonine protein kinase; stevor; telomere: R-CG7; var; varC
            pseudogene.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 108908)
  AUTHORS   Bowman,S., Lawson,D., Basham,D., Brown,D., Chillingworth,T.,
            Churcher,C.M., Craig,A., Davies,R.M, Devlin,K., Feltwell,T.,
            Gentles,S., Gwilliam,R., Hamlin,N., Harris,D., Holroyd,S.,
            Hornsby,T., Horrocks,P., Jagels,K., Jassal,B., Kyes,S., McLean,J.,
            Moule,S., Mungall,K., Murphy,L., Oliver,K., Quail,M.A.,
            Rajandream,M.-A., Rutter,S., Skelton,J., Squares,R., Squares,S.,
            Sulston,J.E., Whitehead,S., Woodward,J.R., Newbold,C. and
            Barrell,B.G.
  TITLE     The complete nucleotide sequence of chromosome 3 of Plasmodium
            falciparum
  JOURNAL   Nature 400 (6744), 532-538 (1999)
  MEDLINE   99376085
REFERENCE   2  (bases 1 to 108908)
  AUTHORS   Bowman,S., Churcher,C., Lawson,D., Quail,M. and Barrell,B.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 108908)
  AUTHORS   Lawson,D., Bowman,S., Rajandream,M.A. and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1998) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On Apr 29, 1999 this sequence version replaced gi:4494019.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This
            sequence is not the entire insert of clone MAL3P8. It may be
            shorter because we only sequence overlapping sections once, or
            longer because we arrange for a small overlap between neighbouring
            submissions.
FEATURES             Location/Qualifiers
     source          1. .108908
                     /organism="Plasmodium falciparum"
                     /strain="3D7"
                     /db_xref="taxon:5833"
                     /chromosome="3"
                     /clone="MAL3P8"
     repeat_region   1. .273
                     /note="PFTelo"
     repeat_region   4530. .6818
                     /note="R-CG7"
     repeat_region   8638. .10564
                     /note="rep11"
     repeat_region   10610. .30944
                     /note="rep20"
     gene            33641. .41158
                     /gene="PFC0005w, MAL3P8.1"
     CDS             join(33641. .38959,39848. .41158)
                     /gene="PFC0005w, MAL3P8.1"
                     /note="PFC0005w (MAL3P8.1), var gene (3D7-varT3-1), len:
                     2209 aa; Similarity to P.falciparum PfEMP-1 gene. Part of
                     a multi-gene family, donor sequence corrected"
                     /codon_start=1
                     /protein_id="CAB39115.2"
                     /db_xref="GI:7264037"
                     /db_xref="SPTREMBL:O97324"
                     /translation="MVRTLDPEEELRGIEDTTAKHIFDRIGKIVHEKAKKNAEQYRSQ
                     LKGSLLKATFEKAPAGQQTPGNTCELKYQWHTNVTKGGNKEYPCRNGTEKRFSEVSGG
                     ECDDSKIKGSNGGACAPFRRLNLCVRNLENINNYGKINNDTLLADVCLAALHEGDSIR
                     GDHDKYKETNDSSQLCTMLARSFADIGDIIRGKDLYRGNNGKDKLEENLKTIFGKIHE
                     GLKNGKTNGIEERYGNDPDFFQLREDWWTANRETVWKAIRCSAPRDADYFIKTVCSGG
                     KTPTQGKCRCIDFSVPTYFDYVPQYLRWFEEWAEDFCRLRKHKLQNAKNKCREKHKDG
                     KKLYCDLNGFDCTQTARGKNKYKYEHDCIECYSSCDHFVHWIDNQKKEFEKQKNKYDK
                     EIKKAYGKNGTTTKETSNGPINNLYVGDFYSKLQQTYGSVDKFLKKLNDEAICKKHPE
                     VEEKTDVNFNENLDDIFSHTKYCQACPLCGLENDSPPWKPKPEKECRDQQIRNFDDNE
                     SNEIKLLVKDKGGQTMVEKLGGLCGNGAKKNNIQEKTWKCYYDKNKENSIGGGDKDYC
                     VLKNDKKNRTQLEIVSFNSLFWRWVTEMLKDSIDWRKEYKNCINNGDKSTCKNVCKKP
                     CDCFQKWGVRKTKEWQQVKAHYEKEDFGKGLTPYKTLEWVLDLSYFPIIKEAHPKEKP
                     VQKMEEIIKKNQENISRVTKQNNSITKFLQQELQEANNCLQKRKQDCKPPQQSAEEGV
                     AKTGQPRAEDEEDSPRPDAGAGEVDDEDDDADNDDEITPRDLNIEVDDLDSKDPEDQV
                     EEEKAKDNTDETVETAKETKEDTDRKGEEKQPKEEVDNVKPCEIVKTLFTNGDNTALN
                     EACKQKYQYGKEKFPNWKCVPTTRGSGEPTGSSGSICVPPRRRRLYVTPLTRLAGGGN
                     TAASQGSGEAAQPVTVTQPQASGGNTQVAVSPGGAASSTSTTESSQLLRQAFIESAAV
                     ETFFLWHKYKVDKEKEDKEKNGRNMVVYTSPVPNDLYEKLKKGEIPDDFLRQMFYTLA
                     DYKDILFSGDKDNKNGYNDILSGDKELHEREKEIKGAIEKHFSNSVKTPSTSGNDAKT
                     WWKENAPHIWEGMICALTYKETSGSGGEKQIEKDSDVYNKFFGSTAATQKGTYESKYK
                     YTDVKLEENSGTGGPRGPNESPNSKPPSSTSENKPTTLDSFIKRPPYFRYLEEWGETF
                     CRERAKRLAQIKVDCRGDENTNRSNDGDGFDCEKKVTNKDVFLEDFNGSSCATCCSSY
                     RKWIERKKIEFEEQENAYEEQQKKNCVNGNNKGGVNGVCGKLEENAAGFLQKLGSCKK
                     DSGEDNGNGNEEDKLNFRQPNVTFRPAENCKPCSLIEIKCKNGVCNGDPTKGECNGET
                     VTAEEIEKMNDLNGNIDMLVSDNGKNEIPEDLKSSCKDANIFKGIREDVWKCGKFRDV
                     DVCVLKNFNKHIHDKKNVLIRTLFKRWLEYFFEDYNRIQKKLKPCIENGKGKEQKCFK
                     GCKENCDCVKKWVEEKEKEWPKIRKRYLEQYKNAGGSDDYKVKSFLEDPQFYNEVNKA
                     VKPCDDLNAFERSIHCNGPNSSQKKDVERRDVVVCLLDKLEKEAKKCEQKHQNSGNPQ
                     QPCEGSTPPDDEEEELLEEDEQNTVGKEKVGNKAPAICGDVEEQKEKEEGDCDKAVTP
                     DSDTGGNGEKEDSRSEEEEEVSGSGDQGSPPAPPPPESPQEKAPAPAPEELPPGPERP
                     PKPAAPTSTPEVPAQPLPSDNTSD1LKTTIPFGIALALTSIALLFLKKKTKSTIDLLR
                     VINIPKSDYGMPTKLSPNRYIPYTSGKYRGKRYIYLEGDSGTDSGYTDHYSDITSSSE
```

```
                SEYEEMDINDIYVPGSPKYKTLIEVVLEPSGNNTTASGKNTPSDTQNDIQNDGIPSDT
                PNTPSDIPKTPSDTPPPITDDEWNTLKDEFISQYLQSEQPNDVPNDYKSGNSSTNTNI
                TTTSRHNVEEKPFIMSIHDRDLYTGEEYNYDMSNNSGIYPSSSNRDSLSGTKVPYSGI
                DLINDSLNSGNQPIDIYDEILKRKENELFGTNHVKQTSIHSVAKLTNSDPIHNQLELF
                HKWLDRHRDMCEKLKNDNERLAKLKEEWENETHSGNIHTSDSNKTLNTDVSIQIDMDD
                GKPKKEFTNMDTILEDLDKYNEPYYDVQDDIYYDVNDHDVSTVDSNAMDIPSKVQIEM
                DVNTKLVKEKYPIADVWDI"
misc_feature    38960. .38965
                /gene="PFC0005w, MAL3P8.1"
                /note="corrected splice donor for PFC0005c"
repeat_region   42416. .43015
                /note="R-FA3"
gene            complement(43045. .44255)
                /gene="PFC0010c, MAL3P8.2"
CDS             complement(join(43045. .43971,44202. .44255))
                /gene="PFC0010c, MAL3P8.2"
                /note="PFC0010c (MAL3P8.2), rifin (3D7-rifT3-1), len: 326
                aa; Similarity to other P.falciparum rifins. Member of a
                multi-gene family; predicted using hexExon"
                /codon_start=1
                /protein_id="CAB39116.1"
                /db_xref="GI:4494021"
                /db_xref="SPTREMBL:O97325"
                /translation="MKIHYINILLFELPLNILIYNQRNHYITRTPKATTRTLCECELY
                APSNYDSDPEMKRVMQQFVDRTTQRFHEYDERMKTTRQKCKEQCDKEIQKIILKDKLE
                KQMAQQLTTLDPNITTEDIPTCICEKSLADKTEKFCLNCGKTMGGVAPGWGLVSGLGY
                VGWTNYVTQTALQKGIEAGVKAGIEELKGFPGLRKLIKVSQIQSFINPANYAEETTYF
                NFIKQVENTKCVVKEHTFCYSTYISKENALSQRASEIAKNAADMAKLAEREVLDKAAP
                VTSGLNTAIIASIVAILVIVLVMIIIYLFLRYRKNKKMKKKLQYIKLLEE"
gene            complement(46448. .47828)
                /gene="PFC0015c"
CDS             complement(46448. .47828)
                /gene="PFC0015c"
                /note="PFC0015c varC pseudogene"
                /codon_start=1
                /pseudo
gene            complement(48956. .49949)
                /gene="PFC0025c, MAL3P8.4"
CDS             complement(join(48956. .49777,49881. .49949))
                /gene="PFC0025c, MAL3P8.4"
                /note="PFC0025c (MAL3P8.4), stevor (3D7-stevorT3-1), len:
                296 aa; Similarity to other P.falciparum stevors. Member
                of a multi-gene family; predicted using hexExon"
                /codon_start=1
                /protein_id="CAB39117.1"
                /db_xref="GI:4494022"
                /translation="MKMYYLKMLLFTFLINTLILPHYENYLNNDYNVSFIQNNTKRTT
                IKSRLLAQTQIHNPHYHNDPELKEIIDKMNEEAIKKYQQTHDPYEQLKEVVEKNGTKY
                SGGNDAEPISTLEKELLETYEEIFGNESDMLKSGMSPNVDEKSSTFECTDINGVKLAK
                TKGRDKYLKHLKGRCTRGIYFCSLGSVILTYIGTTAAKGAAISTAGAGSKNCIYAISI
                FHMLNHKSMSSALTALGTVGCAADYASGATSASAIFLPCGIAALVLLILTVLLIILYI
                WLYRRRKRSWKHECKKHLCK"
gene            complement(52066. .53260)
                /gene="PFC0030c, MAL3P8.5"
CDS             complement(join(52066. .53076,53192. .53260))
                /gene="PFC0030c, MAL3P8.5"
                /note="PFC0030c (MAL3P8.5), rifin, (3D7-rifT3-2), len: 359
                aa; Similarity to other P.falciparum rifins. Member of a
                multi-gene family; predicted using hexExon"
                /codon_start=1
                /protein_id="CAB39118.1"
                /db_xref="GI:4494023"
                /db_xref="SPTREMBL:O97326"
                /translation="MKLLYSNILLFALALNILLTSYYAHNKNKPSITPHHTRTTTSRL
                LSECDTESSIYDSDEEINSVKEIFERQTSRRFEEYEERMITQRQKYKEQRDKDIQKII
                HKDKMEKNLAEKIEKGCLRCGCGLGSVAGSIGLFGAVAINIWKPAALDAAIAKAITEG
                TANISAAGVKAGEATGKVLVISGLKKMGISTLGGKNLESYFATTSYKKVASIAQAVYE
                QHFEKCEFGYLKEGFTPVGDPSRDIHFCQSVWQQTSVVSKTGHYISPKDVIKRTVETM
                VSKAEGPANAAAEFVKATETATIKAAEEKTIETASTQLYSAIGYSILAILIIVLIMLI
                IYLILRYRRKKKMKKKAQYTKLLNE"
gene            58103. .59390
                /gene="PFC0035w, MAL3P8.6"
CDS             join(58103. .58174,58338. .59390)
                /gene="PFC0035w, MAL3P8.6"
                /note="PFC0035w (MAL3P8.6), rifin, (3D7-rifT3-3), len: 375
                aa; Similarity to other P.falciparum rifins. Member of a
                multi-gene family; predicted using hexExon"
                /codon_start=1
                /protein_id="CAB39119.1"
                /db_xref="GI:4494024"
                /db_xref="SPTREMBL:O97327"
                /translation="MRITIKMKVHYINILLLALSLNILANTHQKPSSTPRHIQTTRLL
                CECELYMHNYDNDPEMKRVMQQFHDRTTQRFQEYDERLQEKRQVCKDTCDKEIEKIIL
                KDKIEKELNEKFATLQTDIHSDAIPTCICKTSMADKVEKTCLKCTQNLGGIVAPSSGV
                LAGIAEGALYAWKPNALQTAIEAALKAAADDILVGGIEAGKKVVIGGLDALEIEKLGI
                GSWEPYFTEGYCINVKSLASIIYEKRQTLCGATKSTLDKATCEQIGISIGTMQHDGTY
                GAPGTTPIETVLNGIVEGTKEAADVAAEAARESATNAIKVQETRLLEAGFNSSISSIN
                ASIIAIVVIILIMVIIYFILRYRRKKKLKKKLQYIKLLEE"
gene            61248. .62459
                /gene="PFC0040w,MAL3P8.7"
CDS             join(61248. .61301,61548. .62459)
                /gene="PFC0040w,MAL3P8.7"
                /note="PFC0040w (MAL3P8.7), rifin, (3D7-rifT3-4), len: 322
                aa; Similarity to other P.falciparum rifins. Member of a
                multi-gene family; predicted signal and transmembrane
                sequence; predicted using hexExon"
                /codon_start=1
                /protein_id="CAB39120.1"
                /db_xref="GI:4494025"
                /db_xref="SPTREMBL:O97328"
                /translation="MKIHYINILLFALPLNILVNNQRNHNNSTYHTSNTKTIKSHRSL
```

```
  Query Match              2.3%;  Score 45;  DB 33;  Length 108908;
  Best Local Similarity   55.4%;  Pred. No. 0.29;
  Matches   87;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy   1814 tcaccttcacgaaacacacaaaaacacagcgaattgtagcacctcaaagattcgatagct 1873
          |||   | | |||  | ||||||  | |   |||  ||       || ||||| | |
Db 101106 TCAAAATAAAGAACGAAACAAAATTAAAAATAATATTAAATATAAAATGATTCAAAAATA 101165

Qy   1874 ttttgtcatagtccttagtcttaactcgtatttatttttcgtacggttgtcgagctcaaaa 1933
          || | | |||    |   | ||  ||  | | ||||   ||| |      ||    | |
Db 101166 TTATATTATAACAATCGTTTTTTTTCTTATGTCTATTAGGGTAAGCGAAAAGACAAGATTA 101225

Qy   1934 ataaaatcaaattaagctaaaaaaaaaaaaaaaaaa 1970
          | ||||  |||  ||   ||||||||||||||||||
Db 101226 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 101262


RESULT   9
AC013583
LOCUS       AC013583   173184 bp    DNA             HTG       16-MAR-2000
DEFINITION  Homo sapiens clone RP11-5C23, WORKING DRAFT SEQUENCE, 18 unordered
            pieces.
ACCESSION   AC013583
VERSION     AC013583.4  GI:7248956
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 173184)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-5C23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 173184)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
```

```
          Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
          Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
          Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
          Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Mar 16, 2000 this sequence version replaced gi:6970296.
          All repeats were identified using RepeatMasker:
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          -------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
          -------------- Project Information
              Center project name: L2847
              Center clone name: 5_C_23
          -------------- Summary Statistics
              Sequencing vector: M13; M77815; 100% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 0.960731
              Consensus quality: 160008 bases at least Q40
              Consensus quality: 166750 bases at least Q30
              Consensus quality: 169341 bases at least Q20
              Insert size: 182000; agarose-fp
              Insert size: 171484; sum-of-contigs
              Quality coverage: 4.2 in Q20 bases; agarose-fp
              Quality coverage: 4.5 in Q20 bases; sum-of-contigs
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 18 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1     1031: contig of 1031 bp in length
          *    1032  1131: gap of      100 bp
          *    1132     3436: contig of 2305 bp in length
          *    3437  3536: gap of      100 bp
          *    3537     5182: contig of 1646 bp in length
          *    5183  5282: gap of      100 bp
          *    5283     8761: contig of 3479 bp in length
          *    8762  8861: gap of      100 bp
          *    8862    14445: contig of 5584 bp in length
          *   14446 14545: gap of      100 bp
          *   14546    20653: contig of 6108 bp in length
          *   20654 20753: gap of      100 bp
          *   20754    26231: contig of 5478 bp in length
          *   26232 26331: gap of      100 bp
          *   26332    34259: contig of 7928 bp in length
          *   34260 34359: gap of      100 bp
          *   34360    40262: contig of 5903 bp in length
          *   40263 40362: gap of      100 bp
          *   40363    46964: contig of 6602 bp in length
          *   46965 47064: gap of      100 bp
          *   47065    56070: contig of 9006 bp in length
          *   56071 56170: gap of      100 bp
          *   56171    71111: contig of 14941 bp in length
          *   71112 71211: gap of      100 bp
          *   71212    80835: contig of 9624 bp in length
          *   80836 80935: gap of      100 bp
          *   80936    95321: contig of 14386 bp in length
          *   95322 95421: gap of      100 bp
          *   95422   108414: contig of 12993 bp in length
          *  108415 108514: gap of      100 bp
          *  108515   122869: contig of 14355 bp in length
          *  122870 122969: gap of      100 bp
          *  122970   137578: contig of 14609 bp in length
          *  137579 137678: gap of      100 bp
          *  137679   173184: contig of 35506 bp in length.
FEATURES             Location/Qualifiers
     source          1. .173184
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-5C23"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1031
                     /note="assembly_fragment"
     misc_feature    1132. .3436
                     /note="assembly_fragment"
     misc_feature    3537. .5182
                     /note="assembly_fragment"
     misc_feature    5283. .8761
                     /note="assembly_fragment"
     misc_feature    8862. .14445
                     /note="assembly_fragment"
     misc_feature    14546. .20653
                     /note="assembly_fragment"
     misc_feature    20754. .26231
                     /note="assembly_fragment"
     misc_feature    26332. .34259
                     /note="assembly_fragment"
     misc_feature    34360. .40262
                     /note="assembly_fragment"
     misc_feature    40363. .46964
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    47065. .56070
                     /note="assembly_fragment"
     misc_feature    56171. .71111
                     /note="assembly_fragment"
     misc_feature    71212. .80835
                     /note="assembly_fragment"
     misc_feature    80936. .95321
                     /note="assembly_fragment"
     misc_feature    95422. .108414
                     /note="assembly_fragment"
     misc_feature    108515. .122869
                     /note="assembly_fragment"
     misc_feature    122970. .137578
                     /note="assembly_fragment"
     misc_feature    137679. .173184
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
BASE COUNT    50006 a  37306 c  35778 g  48389 t   1705 others
ORIGIN


  Query Match             2.3%;  Score 44.6;  DB 60;  Length 173184;
  Best Local Similarity   49.0%;  Pred. No. 0.42;
  Matches  119;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy   1678 gttcccttctctttgcactggccctgcttaaggtggcctggttgagttttctgcgacccc 1737
          || ||||| |||      ||  |      || |  | ||   |   |||||       |
Db   5913 GTGCCCTTTTCTCAAAGGTGTGCACAGAAAAAGATGACTAAATACTTTTTCAAAAGTATC 5972

Qy   1738 tgctttaagaactgaaatggaaaactgaaatggatcctgggagttcaactccctgcaaag 1797
            || | || ||  |||||  |  |||  |||    |||  || |||| |        |
Db   5973 ACCTATGAGCACCAAAATGACAGCCTGGGATGAGGACTGTCAGGTCAAATAAGATGGGAC 6032

Qy   1798 acgctagactgctatttcaccttcacgaaacacacaaaaacacagcgaattgtagcacct 1857
          |  ||| | |||  ||||||||||  |    ||| ||     | | |||| |  ||| |
Db   6033 ATCCTACAGTGCACTTTCACCTTCCTAAGGACCACTAATTTTAAACAAATTATTCCACGT 6092

Qy   1858 caaagattcgatagctttttgtcatagtccttagtcttaactcgtatttatttttcgtacg 1917
            |    |   |  ||||     |||    || |  ||   ||  ||   ||||  || |
Db   6093 AATTTGTAAAACGACTTTAAACAATAACAATTTGATTTTTTTCTAATAGTTTTTATTATG 6152

Qy   1918 gtt 1920
            ||
```

```
Db   6153 ATT 6155


RESULT  10
AL136310/c
LOCUS       AL136310   237302 bp    DNA             HTG       20-APR-2000
DEFINITION  Homo sapiens chromosome 6 clone RP3-415N12, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL136310
VERSION     AL136310.5  GI:7634172
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 237302)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 21, 2000 this sequence version replaced gi:7378514.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: dJ415N12
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 193992 bases at least Q40
            Consensus quality: 210387 bases at least Q30
            Consensus quality: 220937 bases at least Q20
            Insert size: 230202; sum-of-contigs
            Insert size: 137822; 12.5% error;'agarose-fp
            Quality coverage: 2.65x in Q20 bases; sum-of-contigs Quality
            coverage: 4.43x in Q20 bases; agarose-fp
            --------------
            * NOTE: This is a 'unfinished' sequence. It currently * consists of
            72 contigs. The true order of the pieces is * not known and their
            order in this sequence record is * arbitrary.  Where the contigs
            adjacent to the vector can * be identified, they are labelled with
            'clone_end' in the * feature table.  Some order and orientation
            information * can tentatively be deduced from paired sequencing
            reads * which have been identified to span the gap between two *
            contigs.  These are labelled as part of the same *
            'fragment_chain', and the order and relative orientation * of the
            pieces within a fragment_chain is reflected in * this file.  Gaps
            between the contigs are represented as * runs of N, but the exact
            sizes of the gaps are unknown. * This record will be updated with
            the finished sequence as * soon as it is available and the
            accession number will be * preserved.
            *       1     3938 contig of 3938 bp in length; fragment_chain 1 *
                4039     7596 contig of 3558 bp in length; fragment_chain 1 *
              7697    25193 contig of 17497 bp in length; fragment_chain 1 *
            25294    48416 contig of 23123 bp in length; fragment_chain 1 *
            48517    79454 contig of 30938 bp in length; fragment_chain 1 *
            79555    88282 contig of 8728 bp in length; fragment_chain 2 *
            88383    92814 contig of 4432 bp in length; fragment_chain 2 *
            92915    94176 contig of 1262 bp in length; fragment_chain 2 *
            94277    96619 contig of 2343 bp in length; fragment_chain 2 *
            96720    98893 contig of 2174 bp in length; fragment_chain 3 *
            98994   100558 contig of 1565 bp in length; fragment_chain 3 *
            100659   102748 contig of 2090 bp in length; fragment_chain 3 *
            102849   104488 contig of 1640 bp in length; fragment_chain 4 *
            104589   105793 contig of 1205 bp in length; fragment_chain 4 *
            105894   109086 contig of 3193 bp in length; fragment_chain 4 *
            109187   110267 contig of 1081 bp in length; fragment_chain 5 *
            110368   111393 contig of 1026 bp in length; fragment_chain 5 *
            111494   112651 contig of 1158 bp in length; fragment_chain 5 *
            112752   114034 contig of 1283 bp in length; fragment_chain 6 *
            114135   115953 contig of 1819 bp in length; fragment_chain 6 *
            116054   117793 contig of 1740 bp in length; fragment_chain 7 *
            117894   119191 contig of 1298 bp in length; fragment_chain 7 *
            119292   120448 contig of 1157 bp in length; fragment_chain 8 *
            120549   122011 contig of 1463 bp in length; fragment_chain 8 *
            122112   123814 contig of 1703 bp in length; fragment_chain 9 *
            123915   125651 contig of 1737 bp in length; fragment_chain 9 *
            125752   126854 contig of 1103 bp in length; fragment_chain 10 *
            126955   129043 contig of 2089 bp in length; fragment_chain 10 *
            129144   130148 contig of 1005 bp in length; fragment_chain 11 *
            130249   131530 contig of 1282 bp in length; fragment_chain 11 *
            131631   134360 contig of 2730 bp in length; fragment_chain 12 *
            134461   139505 contig of 5045 bp in length; fragment_chain 12 *
            139606   141471 contig of 1866 bp in length; fragment_chain 13 *
            141572   143919 contig of 2348 bp in length; fragment_chain 13 *
            144020   145706 contig of 1687 bp in length; fragment_chain 14 *
            145807   149597 contig of 3791 bp in length; fragment_chain 14 *
            149698   152091 contig of 2394 bp in length; fragment_chain 15 *
            152192   153223 contig of 1032 bp in length; fragment_chain 15 *
            153324   154577 contig of 1254 bp in length
            *   154678   156040 contig of 1363 bp in length
            *   156141   157669 contig of 1529 bp in length
            *   157770   158989 contig of 1220 bp in length
            *   159090   169641 contig of 10552 bp in length
            *   169742   171499 contig of 1758 bp in length
            *   171600   172750 contig of 1151 bp in length
            *   172851   174805 contig of 1955 bp in length
            *   174906   177316 contig of 2411 bp in length
            *   177417   178643 contig of 1227 bp in length
            *   178744   182008 contig of 3265 bp in length
            *   182109   184328 contig of 2220 bp in length
            *   184429   186132 contig of 1704 bp in length
            *   186233   190452 contig of 4220 bp in length
            *   190553   194819 contig of 4267 bp in length
            *   194920   197185 contig of 2266 bp in length
            *   197286   198667 contig of 1382 bp in length
            *   198768   201386 contig of 2619 bp in length
            *   201487   202629 contig of 1143 bp in length
            *   202730   203933 contig of 1204 bp in length
            *   204034   206262 contig of 2229 bp in length
            *   206363   209561 contig of 3199 bp in length
            *   209662   212772 contig of 3111 bp in length
            *   212873   214378 contig of 1506 bp in length
            *   214479   218263 contig of 3785 bp in length
            *   218364   221064 contig of 2701 bp in length
            *   221165   222898 contig of 1734 bp in length
            *   222999   224757 contig of 1759 bp in length
            *   224858   226013 contig of 1156 bp in length
            *   226114   228699 contig of 2586 bp in length
            *   228800   230694 contig of 1895 bp in length
            *   230795   233315 contig of 2521 bp in length
            *   233416   234443 contig of 1028 bp in length
            *   234544   237302 contig of 2759 bp in length.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .237302
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP3-415N12"
                     /clone_lib="RPCI-3"
     misc_feature    1. .3938
                     /note="assembly_fragment:01023
                     fragment_chain:1"
     misc_feature    4039. .7596
                     /note="assembly_fragment:01368
                     fragment_chain:1"
     misc_feature    7697. .25193
                     /note="assembly_fragment:00456
```

```
                fragment_chain:1"
    misc_feature    25294. .48416
                    /note="assembly_fragment:01165
                    fragment_chain:1"
    misc_feature    48517. .79454
                    /note="assembly_fragment:00487
                    fragment_chain:1"
    misc_feature    79555. .88282
                    /note="assembly_fragment:00635
                    fragment_chain:2"
    misc_feature    88383. .92814
                    /note="assembly_fragment:00698
                    fragment_chain:2"
    misc_feature    92915. .94176
                    /note="assembly_fragment:00324
                    fragment_chain:2"
    misc_feature    94277. .96619
                    /note="assembly_fragment:00747
                    fragment_chain:2"
    misc_feature    96720. .98893
                    /note="assembly_fragment:00145
                    fragment_chain:3"
    misc_feature    98994. .100558
                    /note="assembly_fragment:01003
                    fragment_chain:3"
    misc_feature    100659. .102748
                    /note="assembly_fragment:00478
                    fragment_chain:3"
    misc_feature    102849. .104488
                    /note="assembly_fragment:01155
                    fragment_chain:4"
    misc_feature    104589. .105793
                    /note="assembly_fragment:01353
                    fragment_chain:4"
    misc_feature    105894. .109086
                    /note="assembly_fragment:00123
                    fragment_chain:4"
    misc_feature    109187. .110267
                    /note="assembly_fragment:01490
                    fragment_chain:5"
    misc_feature    110368. .111393
                    /note="assembly_fragment:00295
                    fragment_chain:5"
    misc_feature    111494. .112651
                    /note="assembly_fragment:01121
                    fragment_chain:5"
    misc_feature    112752. .114034
                    /note="assembly_fragment:00195
                    fragment_chain:6"
    misc_feature    114135. .115953
                    /note="assembly_fragment:00664
                    fragment_chain:6"
    misc_feature    116054. .117793
                    /note="assembly_fragment:00365
                    fragment_chain:7"
    misc_feature    117894. .119191
                    /note="assembly_fragment:01325
                    fragment_chain:7"
    misc_feature    119292. .120448

  Query Match             2.3%;  Score 44.6;  DB 40;  Length 237302;
  Best Local Similarity   49.0%;  Pred. No. 0.45;
  Matches  119;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy   1678 gttcccttctctttgcactggccctgcttaaggtggcctggttgagttttctgcgacccc 1737
          || ||||| |||      ||  | |    || |  | ||   |   |||||          |
Db  91431 GTGCCCTTTTCTCAAAGGTGTGCATAGAAAAAGATGACTAAATACTTTTTCAAAAGTATC 91372

Qy   1738 tgctttaagaactgaaatggaaaactgaaatggatcctgggagttcaactccctgcaaag 1797
            || | || ||  |||||  |  |||  |||    |||  || |||| |        |
Db  91371 ACCTATGAGCACCAAAATGACAGCCTGGGATGAGGACTGTCAGGTCAAATAAGATGGGAC 91312

Qy   1798 acgctagactgctatttcaccttcacgaaacacacaaaaacacagcgaattgtagcacct 1857
```

```
          |  ||| | |||  |||||||| ||   |    ||| ||     | | |||| |  ||| |
Db  91311 ATCCTACAGTGCACTTTCACCATCCTAAGGACCACTAATTTTAAACAAATTATTCCACGT 91252

Qy   1858 caaagattcgatagctttttgtcatagtccttagtcttaactcgtatttatttttcgtacg 1917
           |    |   |  ||||     |||    || | ||  ||  ||  ||   ||||  || |
Db  91251 AATTTGTAAAACGACTTTAAACAATAACAATTTGATTTTTTTCTAATAGTTTTTATTATG 91192

Qy   1918 gtt 1920
           ||
Db  91191 ATT 91189


RESULT  11
DMTNF14B
LOCUS       DMTNF14B      268 bp    DNA             INV       13-NOV-1996
DEFINITION  Drosophila F family transposable element F14 3' region.
ACCESSION   X01937
VERSION     X01937.1  GI:8707
KEYWORDS    transposon.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 268)
  AUTHORS   Di Nocera,P.P., Digan,M.E. and Dawid,I.B.
  TITLE     A family of oligo-adenylate-terminated transposable sequences in
            Drosophila melanogaster
  JOURNAL   J. Mol. Biol. 168 (4), 715-727 (1983)
  MEDLINE   83294542
COMMENT     Data kindly reviewed (18-NOV-1985) by I.B. Dawid.
FEATURES             Location/Qualifiers
     source          1. .268
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
     source          1. .258
                     /organism="Drosophila melanogaster"
                     /transposon="transposable element F1"
                     /db_xref="taxon:7227"
     gene            1. .268
                     /gene="F-element"
                     /note="F element"
                     /allele=""
                     /db_xref="FlyBase:FBgn0000652"
     polyA_signal    214. .219
                     /gene="F-element"
                     /db_xref="FlyBase:FBgn0000652"
     polyA_signal    218. .223
                     /gene="F-element"
                     /db_xref="FlyBase:FBgn0000652"
     repeat_region   259. .268
                     /note="direct repeat (target site duplication)"
BASE COUNT      113 a     52 c     41 g     60 t      2 others
ORIGIN


  Query Match             2.2%;  Score 44.2;  DB 33;  Length 268;
  Best Local Similarity   61.9%;  Pred. No. 0.15;
  Matches   70;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy   1858 caaagattcgatagctttttgtcatagtccttagtcttaactcgtatttatttttcgtacg 1917
           |||   ||| | |  |  |||| |||||    | | ||||      | |   ||| ||
Db    143 CAACTGTTCAAAAAATACTTGTTATAGTTAAGATTTTTAAACTTATTGTTAGTTCTTATA 202

Qy   1918 gttgtcgagctcaaaaataaaatcaaattaagctaaaaaaaaaaaaaaaaaaaa 1970
           |  ||    | |||||||| |||| ||| ||||||||||||||||||||
Db    203 CAAGAAGATTCGATAAATAAAAGCAAAGTAAAATAAAAAAAAAAAAAAAAAA 255


RESULT  12
AE003083/c
LOCUS       AE003083    17196 bp    DNA             1NV       22-MAR-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013385449, complete
```

```
          sequence.
ACCESSION   AE003083
VERSION     AE003083.1  GI:7289169
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 17196)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 17196)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
FEATURES             Location/Qualifiers
     source          1. .17196
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     5642 a   3215 c   3276 g   5063 t
ORIGIN


  Query Match             2.2%;  Score 44.2;  DB 34;  Length 17196;
  Best Local Similarity   61.9%;  Pred. No. 0.34;
  Matches   70;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy    1858 caaagattcgatagctttttgtcatagtccttagtcttaactcgtatttatttttcgtacg 1917
           |||   ||| | |  |  |||| |||||    | | ||||      | |   ||| ||
Db    3677 CAACTGTTCAAAAAATACTTGTTATAGTTAAGATTTTTAAACTTATTGTTAGTTCTTATA 3618

Qy    1918 gttgtcgagctcaaaaataaaatcaaattaagctaaaaaaaaaaaaaaaaaa 1970
           |  ||    | |||||||| |||| |||  ||||||||||||||||||||
Db    3617 CAAGAAGATTCAATAAATAAAAGCAAAGTAAAATAAAAAAAAAAAAAAAAAA 3565


RESULT  13
AE003098/c
LOCUS       AE003098    24781 bp    DNA             INV       22-MAR-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013385318, complete
            sequence.
ACCESSION   AE003098
VERSION     AE003098.1  GI:7289017
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 24781)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 24781)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
FEATURES             Location/Qualifiers
     source          1. .24781
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
                     /map="42B"
BASE COUNT     4986 a   3363 c   3312 g   5751 t   7369 others
ORIGIN
```

```
  Query Match             2.2%;  Score 44.2;  DB 34;  Length 24781;
  Best Local Similarity   61.9%;  Pred. No. 0.37;
  Matches   70;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy   1858 caaagattcgatagctttttgtcatagtccttagtcttaactcgtatttatttttcgtacg 1917
          |||   ||| | |  |  |||| |||||    | | ||||      | |   ||| ||
Db   5222 CAACTGTTCAAAAAATACTTGTTATAGTTAAGATTTTTAAACTTATTGTTAGTTCTTATC 5163

Qy   1918 gttgtcgagctcaaaaataaaatcaaattaagctaaaaaaaaaaaaaaaaaaa 1970
            |  ||    | |||||||| |||| |||  ||||||||||||||||||||
Db   5162 CAAGAAGATTCAATAAATAAAAGCAAAGTAAAATAAAAAAAAAAAAAAAAAA 5110


RESULT  14
AC006589/c
LOCUS       AC006589    51236 bp    DNA             HTG       02-AUG-1999
DEFINITION  Drosophila melanogaster chromosome 3 clone DS02393 (D521) map
            91F10-91F11 strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 38
            unordered pieces.
ACCESSION   AC006589
VERSION     AC006589.3  GI:5670668
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 51236)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 51236)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Aug 2, 1999 this sequence version replaced gi:5630006.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases. P1 library location:
            25-89.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 38 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      645: contig of 645 bp in length
            *      646      725: gap of unknown length
            *      726     1453: contig of 728 bp in length
            *     1454     1533: gap of unknown length
            *     1534     2505: contig of 972 bp in length
            *     2506     2585: gap of unknown length
            *     2586     3383: contig of 798 bp in length
            *     3384     3463: gap of unknown length
            *     3464     4082: contig of 619 bp in length
            *     4083     4162: gap of unknown length
            *     4163     4685: contig of 523 bp in length
            *     4686     4765: gap of unknown length
            *     4766     5114: contig of 349 bp in length
            *     5115     5194: gap of unknown length
            *     5195     5855: contig of 661 bp in length
            *     5856     5935: gap of unknown length
            *     5936     6475: contig of 540 bp in length
            *     6476     6555: gap of unknown length
            *     6556     7248: contig of 693 bp in length
            *     7249     7328: gap of unknown length
            *     7329     7940: contig of 612 bp in length
            *     7941     8020: gap of unknown length
            *     8021     8641: contig of 621 bp in length
            *     8642     8721: gap of unknown length
            *     8722    10170: contig of 1449 bp in length
            *    10171    10250: gap of unknown length
            *    10251    11384: contig of 1134 bp in length
            *    11385    11464: gap of unknown length
            *    11465    13132: contig of 1668 bp in length
            *    13133    13212: gap of unknown length
            *    13213    13972: contig of 760 bp in length
            *    13973    14052: gap of unknown length
            *    14053    14976: contig of 924 bp in length
            *    14977    15056: gap of unknown length
            *    15057    16235: contig of 1179 bp in length
            *    16236    16315: gap of unknown length
            *    16316    17415: contig of 1100 bp in length
            *    17416    17495: gap of unknown length
            *    17496    19141: contig of 1646 bp in length
            *    19142    19221: gap of unknown length
            *    19222    21340: contig of 2119 bp in length
            *    21341    21420: gap of unknown length
            *    21421    23484: contig of 2064 bp in length
            *    23485    23564: gap of unknown length
            *    23565    26380: contig of 2816 bp in length
            *    26381    26460: gap of unknown length
            *    26461    28748: contig of 2288 bp in length
            *    28749    28828: gap of unknown length
            *    28829    31555: contig of 2727 bp in length
            *    31556    31635: gap of unknown length
            *    31636    33537: contig of 1902 bp in length
            *    33538    33617: gap of unknown length
            *    33618    35699: contig of 2082 bp in length
            *    35700    35779: gap of unknown length
            *    35780    38419: contig of 2640 bp in length
            *    38420    38499: gap of unknown length
            *    38500    41615: contig of 3116 bp in length
            *    41616    41695: gap of unknown length
            *    41696    45871: contig of 4176 bp in length
            *    45872    45951: gap of unknown length
            *    45952    46570: contig of 619 bp in length
            *    46571    46650: gap of unknown length
            *    46651    47207: contig of 557 bp in length
            *    47208    47287: gap of unknown length
            *    47288    47762: contig of 475 bp in length
            *    47763    47842: gap of unknown length
            *    47843    48504: contig of 662 bp in length
            *    48505    48584: gap of unknown length
            *    48585    49182: contig of 598 bp in length
            *    49183    49262: gap of unknown length
            *    49263    49883: contig of 621 bp in length
            *    49884    49963: gap of unknown length
            *    49964    50519: contig of 556 bp in length
            *    50520    50599: gap of unknown length
            *    50600    51236: contig of 637 bp in length.
FEATURES             Location/Qualifiers
     source          1. .51236
                     /organism="Drosophila melanogaster"
```

```
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="3"
                     /map="91F10-91F11"
                     /clone="DS02393 (D521)"
                     /clone_lib="P1 library, partial Sau3A in pAd10sacBII"
BASE COUNT    13212 a  10748 c  10575 g  13740 t   2961 others
ORIGIN


  Query Match             2.2%;  Score 44.2;  DB 41;  Length 51236;
  Best Local Similarity   61.9%;  Pred. No. 0.43;
  Matches   70;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy   1858 caaagattcgatagctttttgtcatagtccttagtcttaactcgtatttatttttcgtacg 1917
           |||   ||| | |  |  |||| |||||    | | ||||      | |   ||| ||
Db  50240 CAACTGTTCAAAAAATACTTGTTATAGTTAAGATTTTTAAACTTATTGTTAGTTCTTATA 50181

Qy   1918 gttgtcgagctcaaaaataaaatcaaattaagctaaaaaaaaaaaaaaaaaaa 1970
             |  ||    | ||||||| |||| |||  ||||||||||||||||||||
Db  50180 CAAGAAGATTCAATAAATAAAAGCAAAGTAAAATAAAAAAAAAAAAAAAAAA 50128


RESULT  15
AC004336/c
LOCUS       AC004336    61267 bp    DNA             INV       29-AUG-1998
DEFINITION  Drosophila melanogaster DNA sequence (P1 DS03544 (D234)), complete
            sequence.
ACCESSION   AC004336
VERSION     AC004336.1  GI:3492863
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (Subclones in sac from P1 clone DS03544
            (D234)) DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 61267)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
            Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
            Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
            Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
            Yee,A., Zhang,R., Zieran,L.L. and Rubin,G.M.
  TITLE     Sequencing of Drosophila chromosome 2R, region 55A2-55B2
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 61267)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
            Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
            Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
            Zieran,L.L. and Kimmel,B.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-1998) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Aug 29, 1998 this sequence version replaced gi:3419813.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://fruitfly.berkeley.edu/sequence/) or send
            email to bdgp@fruitfly.berkeley.edu.
            Library location: 88-37.
FEATURES             Location/Qualifiers
     source          1. .61267
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
                     /map="55A2-55B2"
                     /clone="P1 DS03544 (D234)"
                     /note="DS03544 (D234) is a bridge overlapping DS08860
                     (D181) and extending to p1 end."
BASE COUNT    17069 a  12971 c  13472 g  17755 t
ORIGIN


  Query Match             2.2%;  Score 44.2;  DB 34;  Length 61267;
  Best Local Similarity   61.9%;  Pred. No. 0.44;
  Matches   70;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy   1858 caaagattcgatagctttttgtcatagtccttagtcttaactcgtatttatttttcgtacg 1917
           |||   ||| | |  |  |||| |||||    | | ||||      | |   ||| ||
Db  20209 CAACTGTTCAAAAAATACTTGTTATAGTTAAGATTTTTAAACTTATTGTTAGTTCTTATA 20150

Qy   1918 gttgtcgagctcaaaaataaaatcaaattaagctaaaaaaaaaaaaaaaaaaa 1970
             |  ||    | ||||||| |||| |||  ||||||||||||||||||||
Db  20149 CAAGAAGATTCAATAAATAAAAGCAAAGTAAAATAAAAAAAAAAAAAAAAAA 20097


Search completed: September  4, 2000, 00:00:24
Job time: 19444 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2000, 18:36:20 ; Search time 11551.9 Seconds
(without alignments)
528.216 Million cell updates/sec

Title: US-09-332-522B-1
Perfect score: 3419
Sequence: 1 ggtttaattacccaagtttg.....taaatacgattatatataaa 3419

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: em_fun:*
17: em_hum1:*
18: em_hum2:*
19: em_in:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_sy:*
29: em_un:*
30: em_vi:*
31: gb_htg1:*
32: gb_htg2:*
33: gb_in1:*
34: gb_in2:*
35: em_ba1:*
36: em_ba2:*
37: em_hum3:*
38: em_hum4:*
39: gb_pr4:*
40: gb_htg3:*
41: gb_htg4:*
42: gb_htg5:*
43: gb_htg6:*
44: gb_htg7:*
45: em_htg1:*
46: em_htg2:*
47: em_htg3:*
48: em_hum5:*
49: gb_pl3:*
50: gb_pr5:*
51: gb_htg8:*
52: gb_htg9:*
53: gb_htg10:*
54: gb_htg11:*
55: gb_htg12:*
56: gb_htg13:*
57: gb_htg14:*
58: gb_in3:*
59: gb_htg15:*
60: gb_htg16:*
61: gb_htg17:*
62: em_htg4:*
63: em_htg5:*
64: em_htg6:*
65: em_htg7:*
66: em_hum6:*
67: gb_htg18:*
68: gb_htg19:*
69: gb_htg20:*
70: gb_htg21:*
71: gb_htg22:*
72: gb_htg23:*
73: gb_htg24:*
74: gb_htg25:*
75: gb_htg26:*
76: gb_htg27:*
77: gb_htg28:*
78: gb_htg29:*
79: gb_htg30:*
80: gb_htg31:*
81: gb_vi1:*
82: gb_vi2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	1008.4	29.5	95968	33	CEY47D3B	AL031635 Caenorhab
c 2	1008.4	29.5	197735	31	CEY66A7	AL022282 Caenorhab
c 3	1008.4	29.5	337565	31	CEY47D3	Z98865 Caenorhabdi
c 4	604.8	17.7	36780	31	CEH10N23	Z94157 Caenorhabdi
5	535	15.6	296699	31	CEY48A6	Z92854 Caenorhabdi
6	424.2	12.4	36780	31	CEH10N23	Z94157 Caenorhabdi
c 7	71.8	2.1	7218	5	I66494	I66494 Sequence 14
8	57.4	1.7	3816	12	CGU09103	U09103 Cricetulus
9	56.8	1.7	1209	12	AB017337	AB017337 Mus muscu
10	55.6	1.6	1767	12	CGU22819	U22819 Cricetulus
11	55.6	1.6	1914	12	CGU12329	U12329 Cricetulus
12	55.6	1.6	4425	12	CGU22818	U22818 Cricetulus
13	55.6	1.6	4987	12	GU12330	U12330 Cricetulus
14	52.6	1.5	3115	12	RATADD1A	L16995 Rat add1 mR
15	51	1.5	4154	5	AR069298	AR069298 Sequence
16	51	1.5	4154	5	I22310	I22310 Sequence 37
17	51	1.5	4154	50	U00968	U00968 Human SREBP
18	49.4	1.4	3768	33	DMU38238	U38238 Drosophila
19	49.4	1.4	44517	42	AC018013	AC018013 Drosophil
c 20	49.4	1.4	273521	34	AE003515	AE003515 Drosophil
c 21	47.8	1.4	92677	51	AC009382	AC009382 Drosophil
22	47.6	1.4	4249	9	HSU02031	U02031 Human stero
23	47.6	1.4	5197	5	AR069307	AR069307 Sequence
24	47.6	1.4	5197	5	I22319	I22319 Sequence 53

	25	46.8	1.4	142947	55	AC010488	AC010488 Homo sapi
c	26	44.6	1.3	127485	50	HUAC002039	AC002039 Homo sapi
	27	43.8	1.3	433	3	AF102873	AF102873 Sus scrof
	28	43.8	1.3	976	7	SOYNOD75	M16976 Soybean ear
	29	43.8	1.3	3060	5	I08177	I08177 Sequence 1
	30	43.8	1.3	3072	7	GMENOD2A	X16875 Soybean ENO
	31	43.8	1.3	4584	7	GMENOD2B	X16876 Soybean ENO
	32	43.8	1.3	184452	40	AL138924	AL138924 Homo sapi
	33	43.6	1.3	156107	78	AC016898	AC016898 Homo sapi
	34	43.6	1.3	176929	39	AC004887	AC004887 Homo sapi
c	35	43	1.3	45993	1	AB012957	AB012957 Vibrio ch
	36	42.6	1.2	29605	7	AB020747	AB020747 Arabidops
	37	42.6	1.2	191375	55	AC023167	AC023167 Mus muscu
	38	42	1.2	40351	11	AC005620	AC005620 Homo sapi
c	39	41.8	1.2	181671	31	AP001024	AP001024 Homo sapi
	40	41.4	1.2	12668	34	AF079445	AF079445 Dictyoste
c	41	41.4	1.2	14363	1	VCOTNREGN	X90547 Vibrio chol
c	42	41.4	1.2	46721	1	AB012956	AB012956 Vibrio ch
c	43	41.4	1.2	90765	73	AC010483	AC010483 Homo sapi
	44	41.4	1.2	98491	12	AF132039	AF132039 Mus muscu
c	45	41.4	1.2	168239	60	AC021367	AC021367 Homo sapi

ALIGNMENTS

RESULT 1
CEY47D3B/c
LOCUS CEY47D3B 95968 bp DNA INV 14-DEC-1999
DEFINITION Caenorhabditis elegans cosmid Y47D3B, complete sequence.
ACCESSION AL031635 Z98865
VERSION AL031635.1 GI:3646936
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 95968)
AUTHORS none.
TITLE Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
MEDLINE 99069613
REMARK The C.elegans Sequencing Consortium.
Erratum:[[published errata appear in Science 1999 Jan 1;283(5398):35 and 1999 Mar 26;283(5410):2103]]
REFERENCE 2 (bases 1 to 95968)
AUTHORS Matthews,L.
TITLE Direct Submission
JOURNAL Submitted (22-SEP-1998) Nematode Sequencing Project, Sanger Centre, Hinxton, Cambridge CB10 1RQ, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.
For a graphical representation of this sequence and its analysis see:-
http://webace.sanger.ac.uk/cgi-bin/display?db=wormace&class=Sequence &object=Y47D3B
Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone Y47D3B. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
The true left end of clone H10N23 is at 61856 in this sequence. The true right end of clone T28D6 is at 104 in this sequence. The true right end of clone Y47D3 is at 95968 in this sequence. The start of this sequence (1. .104) overlaps with the end of sequence Z81134. The end of this sequence (95863. .95968) overlaps with the start of sequence AL032622.

FEATURES Location/Qualifiers
source 1. .95968
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="III"
/clone="Y47D3B"
gene 7188. .12736
/gene="Y47D3B.1"
CDS join(7188. .7236,7287. .7436,7488. .7605,7654. .7713, 8884. .9277,10132. .10318,11510. .11678,11731. .11860, 12379. .12481,12657. .12736)
/gene="Y47D3B.1"
/codon_start=1
/protein_id="CAA21039.1"
/db_xref="GI:3881005"
/db_xref="SPTREMBL:Q9XX03"
/translation="MWAMKNIGDTFLRKDINTRQIILGSFWRDEREQNVDGRFAVIHF LGDGRIEHPLYCFSIQANNRYLITKATVQRIHRGKRASSDICSWAGHIAECDLDVHAL KEFKISTTPNVEDAMLITPELPLREPRHDMVVCMAPMYIYTDWEILVTGIELWLAMGA TKIVVPVQSASNATYRILQEYERKGLVIIRTWPKWPIMSDTNPNGLVLSRGIEESHVN CLFFVKPWSDIVVFSDIDDFLLPLDPSTISPGDNLQILKNIFAEHPQAGSLLFEHRDV QFVPPNRQGDQSLTNFNFEFLRSSKNKMNCNVWRMKTRVVVNASRVDSVNMHETGIHR FGYVQTRVPCRQAHFYHLRHSHNTVPSPTPINMSPLADMLNKQWQTRVEGFTGMKDEV LNKTNTEYLEDFDRCMGAINDEHWTMKVSRCLTPHVCFSRLRREVDCVAVASEYEFYR SGKVRIQKAEVNCESPVPPFTSGNHYFAP"
gene complement(12895. .17349)
/gene="Y47D3B.2a"
CDS complement(join(12895. .12950,13560. .13955,17307. .17349))
/gene="Y47D3B.2a"
/note="cDNA EST yk265b3.5 comes from this gene; cDNA EST EMBL:C10452 comes from this gene; cDNA EST EMBL:C12093 comes from this gene; cDNA EST EMBL:D70235 comes from this gene; cDNA EST yk552e6.3 comes from this gene"
/codon_start=1
/protein_id="CAA21040.1"
/db_xref="GI:3881006"
/db_xref="SPTREMBL:Q9XX02"
/translation="MRNSLFTTLFFGLAALVMVLNAQYTSELEEDEKRGGARAMLHKR GGARAFSADVGDDYKRGGARAFYDEKRGGARAFLTEMKRGGARVFQGFEDEKRGGAKA FMMDKRGGGRAFGDMMKRGGARAFVENSKRDEDWVIRPFEDDRLEKRGGGRSFPVKPG RLDD"
gene complement(13549. .17349)
/gene="Y47D3B.2b"
CDS complement(join(13549. .13955,17307. .17349))
/gene="Y47D3B.2b"
/codon_start=1
/protein_id="CAB54383.1"
/db_xref="GI:5824792"
/translation="MRNSLFTTLFFGLAALVMVLNAQYTSELEEDEKRGGARAMLHKR GGARAFSADVGDDYKRGGARAFYDEKRGGARAFLTEMKRGGARVFQGFEDEKRGGAKA FMMDKRGGGRAFGDMMKRGGARAFVENSKRDEDWVIRPFEDDRLGVF"
gene 20481. .30076
/gene="Y47D3B.3"
CDS join(20481. .20514,21742. .22022,22112. .22190,23753. .23852, 24487. .24611,24655. .24756,25563. .25693,25796. .25891, 27970. .28341,29197. .29332,29684. .29742,29801. .30076)
/gene="Y47D3B.3"
/note="predicted using Genefinder"
/codon_start=1
/protein_id="CAA21038.1"
/db_xref="GI:3881004"
/db_xref="SPTREMBL:Q9XX04"
/translation="MKNKILKVRSRGLFIFGQNFLISKYLRTCISLLDPQLIDLKLQT RMVPVPATFAIIVSFFNFSLLLFNHTYISHYATLSLSSFSFFFRHIFPHLPQQSTSFD QEVFLIYSVLLLVIVRSSIAKDYDIEILPENCDCLDWYGACKVNGKKWVDDDVWKYSC DGKERSDATFTGCQGNGGDVAVGQNTTIEDLWQVCTQDAQKLHYEIEPFCEFNGTKIR VGKEYQDGSFQWLCLSTGRWITGCYYFNETHTDLLLRVDEHAYNGLIEHVCSKRQEYP

```
                AIVQYYTQVRKDVDVKHPTNKGVNRNFPEPLQKLIDEDINTIFLSFCFVSCNTTETNS
                TTTPIPHSTSEAPLKCTRDYHCPSSCDVCRSGICTHILNLSCSRDYHCPACEVCRSGK
                CIGGEEQKKCETENDCEAGGWCMEGRCVENKCDEDSDCAVERAICKVGIEQIFFYFVS
                LNFNEKYWGEHAQQSNFVEFCRITPAMNVSTFILKILLIFIFIFWKFKETKFGEENSP
                DYEDNFDAEKDEIFVKNYLKLMQMGRAPARPVVEKKHNVEESYWCFLSPPKNVRKPRS
                WLKMAGIEVYSAYFDDRENSLFPGNSAIQV"
gene            30854. .32478
                /gene="Y47D3B.4"
CDS             join(30854. .30889,31202. .31345,31929. .32127,32258. .32478)
                /gene="Y47D3B.4"
                /note="predicted using Genefinder"
                /codon_start=1
                /protein_id="CAA21037.1"
                /db_xref="GI:3881003"
                /db_xref="SPTREMBL:Q9XX05"
                /translation="MPMCLHLAFQLLEDISGRLLEWIEMQYLFGADSVSIYHFYLTNQ
                THKVLDYYEKHTNLKVIPLSLPSGNPNSPIERSKYLKENRPQKRRHELLPYNDCFYNY
                IHTHRYVLTIDIDEVIVPIQFDNYGDLLRDFEAKSFGYRLSSISSRNVFKFPSNFTIF
                PRHQYMISNRKRSRKTSTKGEYGKSFSRSVRELFCLIFA"
gene            34913. .48475
                /gene="Y47D3B.5"
CDS             join(34913. .35016,38489. .38669,39530. .39640,39697. .39824,
                39984. .40076,40805. .40983,41034. .41172,42063. .42283,
                42356. .42498,42550. .42774,43192. .43259,43308. .43485,
                43538. .43666,44148. .44300,44351. .44419,44474. .44658,
                44702. .44801,45577. .45675,45727. .46265,47093. .47246,
                47606. .47707,48365. .48475)
                /gene="Y47D3B.5"
                /note="cDNA EST yk449e10.5 comes from this gene; cDNA EST
                EMBL:D35319 comes from this gene; cDNA EST EMBL:C12322
                comes from this gene; cDNA EST EMBL:D32633 comes from this
                gene"
                /codon_start=1
                /protein_id="CAA21041.1"
                /db_xref="GI:3881007"
                /db_xref="SPTREMBL:Q9XX01"
                /translation="MPNPSPTPLQASPTTTLLADGYTDDALPSKYDNFRASRPYSRES
                LLSNTSLGSKVRFSQFSFQSRDDLTTIDDEEEEEYRKPAPPEPTKFLQVSRTTVRMLA
                GGFSRAGSLEGVETGGIVGPPMSPISFLGKCKHYYDKYKLDRISASVLLVLYSFLGAW
                VFYLFEHDYEREVKLKERIDLRMLRNDTFQRISSMVFRQQGLANFEEVFIDYEKKLHV
                VRLPECLDWDYWGALFYVGTLFTTIGYGNIYPRTALGRAASVVYAIVGIPLVLAILSK
                CGKWMTDSLSEKWQQHRIQITEKAKMTKNRLRGKKILKSGEIVEANTGAEADPEKKPE
                VESRTIPIWLALLICVVYVAGCSSLFLLWETRWTFFTSLYFFCISLLTIGLGDIVPDK
                PHMFIVMFVLVIVGLSIVSMFISVVQIKIEEWLCRMIEQIKEEYMKKDGLVDPEEIKR
                IFANDPVLSFVAPKIMSGEQNHMLKDTIKKFDRLLESKEIQTDFPVKGMATQVEKYEQ
                SMACDPMSNHLKPAEHQGTQWSRVVDETGEGPRNVEELKLQRDDATSISRSNDFRDSI
                SDATSLPMDSMSSPVAKKRNKRNTECAVCQCDLPMGNTAEFKADGTAQTDLAQFQIDE
                IAIKLANLQTQRVRPPIVEMATSAFMEDSPLNSEALDLPRGLSQADINVICEAMSSSF
                FKNSQEALKQIADIGVGKDYETTFSDKSQLTSLSLPTVPTRNSIATSPVVSLCEDLKA
                SKKSKKSRSISSSPIPPNSLHEEETQTPLRQNMVHMATDPCTSQLVHRSTVTSPVGRE
                DAETEMPQKMMDRMVSPVSTAIRDQNTSPVEEQTPDFDDRSIQTSISEWFGKLLKKKD
                ESQQTSICDDVVGKSKKEGKKKKASRDENSQASSLGSDVFTSSNPSSTSDDSEKNHKK
                IKKTLIAPNMTASVSMSTQYSPPPGFDRLSEPHSTRARSSSTSGMGTSIFEDETRQEV
                IVQTDDSYLKIARRLDEYRNNKTQFLPVCAAAPLSSKEVEPFKSDRPSERSHYYFGGG
                RRASLGRGRKKPRQDMRYGESQTGQSMDAELLKEVLSSEKSRSISPKRPCNKRGVLSR
                HPSLPTGISRGKVSDYVIQHEKGIHNPATNRQSPIKIIRQYSLDHDVI"
gene            51392. .52342
                /gene="Y47D3B.6"
CDS             join(51392. .51449,51544. .51641,51689. .52342)
                /gene="Y47D3B.6"
                /note="predicted using Genefinder; cDNA EST yk506e5.3
                comes from this gene"
                /codon_start=1
                /protein_id="CAA21043.2"
                /db_xref="GI:5824793"
                /translation="MFRLTICATLLAFTAAQVPVAPLYSACSAAVPCAAPAVCQASVC
                IPTAAAPPMMMAPPAPLPAYAPAPLPAYAPAPLPAPVFAAPAPAPVFAAPFAAPPAPV
                IAAQPALGCAAPCLGAAQCVAGACACANPAVVYSPVIGCQPAIPVAPAPVVAGRLIPQ
                ALPGSPCEPGVECTGGSVCSVGVCLCPPELIQEGTVCVARTIYGVVPPPVIPVAPVIP
                VALGAPCAAPVAPVYASAAPACVPGAVCSAGVCQCAGAYEPLQGACVRRRL"
gene            complement(52583. .66730)
                /gene="Y47D3B.7"
CDS             complement(join(52583. .52774,52830. .53190,61122. .61275,
                61322. .61549,61609. .62005,62315. .62843,62889. .63082,
                63606. .63841,64186. .64605,64659. .64767,65539. .65848,
                66107. .66239,66652. .66730))
                /gene="Y47D3B.7"
                /note="predicted using Genefinder; similar to
                Helix-loop-helix DNA-binding domain; cDNA EST EMBL:D34815
                comes from this gene; cDNA EST EMBL:D35004 comes from this
                gene; cDNA EST EMBL:D68853 comes from this gene; cDNA EST
                EMBL:D67853 comes from this gene; cDNA EST yk349h3.5 comes
```

```
  Query Match             29.5%;  Score 1008.4;  DB 33;  Length 95968;
  Best Local Similarity   76.6%;  Pred. No. 1.6e-280;
  Matches 1508;  Conservative    0;  Mismatches    1;  Indels  459;  Gaps    4;

Qy    1304 attaccgtacccggaaccaattcaatacactgaatactctgcccgatcacccgtcgaatc 1363
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63089 ATTACAGTACCCGGAACCAATTCAATACACTGAATACTCTGCCCGATCACCCGTCGAATC 63030

Qy    1364 atctccttctccacctagaaatgagagaaaacgatcacgaatgagcacaacgactcctat 1423
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63029 ATCTCCTTCTCCACCTAGAAATGAGAGAAAACGATCACGAATGAGCACAACGACTCCTAT 62970

Qy    1424 gaagaatggaactagagatggatcttcgaaagttaccctttttgcgatgctcctagcagt 1483
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62969 GAAGAATGGAACTAGAGATGGATCTTCGAAAGTTACCCTTTTTGCGATGCTCCTAGCAGT 62910

Qy    1484 tctgatttttaatccgattg---------------------------------------- 1503
           ||||||||||||||||||||
Db   62909 TCTGATTTTTAATCCGATTGGGTAAGCTTCTATTCATCTGAAAAATAGCGTTAAATCCTT 62850

Qy    1504 -----gattgctcgctggaagtgcgatattctcaaaagccgctgcagaagctccgattgc 1558
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62849 GTTTAGATTGCTCGCTGGAAGTGCGATATTCTCAAAAGCCGCTGCAGAAGCTCCGATTGC 62790

Qy    1559 ctccccgttcgagcatggaagagtgattgatgacccggatggaactagcactcggacgct 1618
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62789 CTCCCCGTTCGAGCATGGAAGAGTGATTGATGACCCGGATGGAACTAGCACTCGGACGCT 62730

Qy    1619 tttctgggaagggagtatcatcaatatgagctatgtctgggtgttcaacatcttaatgat 1678
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62729 TTTCTGGGAAGGGAGTATCATCAATATGAGCTATGTCTGGGTGTTCAACATCTTAATGAT 62670

Qy    1679 catatatgtggttgtcaaactgctgatccatggtgaccctgttcaagacttcatgtccgt 1738
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62669 CATATATGTGGTTGTCAAACTGCTGATCCATGGTGACCCTGTTCAAGACTTCATGTCCGT 62610

Qy    1739 ttcatggcagacttttgtgacgactcgagagaaggcgagagccgagttgaactctggaaa 1798
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62609 TTCATGGCAGACTTTTGTGACGACTCGAGAGAAGGCGAGAGCCGAGTTGAACTCTGGAAA 62550

Qy    1799 tttgaaagatgctcagagaaagttctgcgagtgtcttgcaacgttggatcgatcgcttcc 1858
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62549 TTTGAAAGATGCTCAGAGAAAGTTCTGCGAGTGTCTTGCAACGTTGGATCGATCGCTTCC 62490

Qy    1859 atcaccgggggttgattcggtgttttcggttggctgggaatgcgttcgacatcttttgaa 1918
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62489 ATCACCGGGGGTTGATTCGGTGTTTTCGGTTGGCTGGGAATGCGTTCGACATCTTTTGAA 62430

Qy    1919 ttggttgtggatcgggagatacatcgcaagaaggcgcaggtccaccacgaagcctgtctc 1978
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62429 TTGGTTGTGGATCGGGAGATACATCGCAAGAAGGCGCAGGTCCACCACGAAGCCTGTCTC 62370

Qy    1979 agtcgtttgtaggagtcatgcgcagactgcagttctctatcatgaaattcat-------- 2030
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62369 AGTCGTTTGTAGGAGTCATGCGCAGACTGCAGTTCTCTATCATGAAATTCATCAGGTTTA 62310

Qy    2031 ---------------------------------------------------------- 2030

Db   62309 CTGTCAAATTCCATAAATTTTGCATTCAAAAATATATTTACTAGAGTAATCCTACAGTAC 62250

Qy    2031 ---------------------------------------------------------- 2030

Db   62249 TGCTACAGTACCCCTACAGTACTACTACAGTACCCTGACCATATCCCCACTAACTCCAAA 62190
```

```
Qy   2031 ------------------------------------------------------------ 2030

Db  62189 CTAATATCCCTTCAAAAGACAAAAAGTCAATTTTTCCAAAACTACAGTAATCCTTCAGTA 62130

Qy   2031 ------------------------------------------------------------ 2030

Db  62129 CTCCTACGGTACCATTACAGTACTACTACAGTACCCTGACCATATCCCCCCACTAACTCC 62070

Qy   2031 ------------------------------------------------------------ 2030

Db  62069 GAACTAATATCCCTTCAAGAGACAGAAACATGTATTTTTTATGTATAAGTAATGATTTTT 62010

Qy   2031 -cagctccatctaatgggtatcactggaaacttcgaagacacctatgaaccatccgccct 2089
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  62009 CCAGCTCCATCTAATGGGTATCACTGGAAACTTCGAAGACACCTATGAACCATCCGCCCT 61950

Qy   2090 aacgggcctcttcatgtccctctgtgcagtaaaccttgctgaagctgccggagcatcaaa 2149
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61949 AACGGGCCTCTTCATGTCCCTCTGTGCAGTAAACCTTGCTGAAGCTGCCGGAGCATCAAA 61890

Qy   2150 cgacggacttccacgcgccgtcatggctcagatctacatttctgcatccatccaatgccg 2209
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61889 CGACGGACTTCCACGCGCCGTCATGGCTCAGATCTACATTTCTGCATCCATCCAATGCCG 61830

Qy   2210 tttggctcttccgaacctactcgcaccattcttctcgggatacttttttacgaagagctcg 2269
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61829 TTTGGCTCTTCCGAACCTACTCGCACCATTCTTCTCGGGATACTTTTTACGAAGAGCTCG 61770

Qy   2270 aaggcacgtgcgtcgagctccggagcactcggtgtcccatttgttatggatcttccatcc 2329
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61769 AAGGCACGTGCGTCGAGCTCCGGAGCACTCGGTGTCCCATTTGTTATGGATCTTCCATCC 61710

Qy   2330 agcgacaagaaagttcatgtcagatgcgaaaaggttggagcatgtgttgagctcgaagca 2389
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61709 AGCGACAAGAAAGTTCATGTCAGATGCGAAAAGGTTGGAGCATGTGTTGAGCTCGAAGCA 61650

Qy   2390 gaagcagttgagatttgggtcttttgtggaagatgagcaat------------------- 2430
           |||||||||||||||||||||||||||||||||||||||
Db  61649 GAAGCAGTTGAGATTTGGGTCTTTTGTGGAAGATGAGCAATGTAAGTATTTTAAAATAAA 61590

Qy   2431 ----------------------------------------tatccccacttgctcgaatc 2450
                                                    ||||||||||||||||||||
Db  61589 ATAAAATAAAATAAAAAATAAATAAACTTTCATTTTCCAGTATCCCCACTTGCTCGAATC 61530

Qy   2451 cgaacaacgctgaaagtgtacctactctccaaacttgtacaggaacttgtcggtggtgac 2510
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61529 CGAACAACGCTGAAAGTGTACCTACTCTCCAAACTTGTACAGGAACTTGTCGGTGGTGAC 61470

Qy   2511 gagatctttacaaaaaatgtggaacgcatcctaaatgacaatgaccgtctcgatgatgaa 2570
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61469 GAGATCTTTACAAAAAATGTGGAACGCATCCTAAATGACAATGACCGTCTCGATGATGAA 61410

Qy   2571 gtagacgtggttgatgtttcaagacttttggtgacaatttcaacgcagtgcgctgccatt 2630
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61409 GTAGACGTGGTTGATGTTTCAAGACTTTTGGTGACAATTTCAACGCAGTGCGCTGCCATT 61350

Qy   2631 ttgactaatgagaaggatgagtcagcga------------------------------ 2658
           |||||||||||||||||||||||||||
Db  61349 TTGACTAATGAGAAGGATGAGTCAGCGAGTTAGTGAACACGAATTTTAAGATTGAAAATG 61290

Qy   2659 --------------aattcggaacctggatctctcgaaacggagatgcttgttgcacatg 2704
                         ||||||||||||||||||||||||||||||||||||||||||||
Db  61289 TTATAAATTTCTAGAATTCGGAACCTGGATCTCTCGAAACGGAGATGCTTGTTGCACATG 61230

Qy   2705 gtggacgcacgttctgacatgtggaatctattggaggagtaacaagaatgagctggcacg 2764
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61229 GTGGACGCACGTTCTGACATGTGGAATCTATTGGAGGAGTAACAAGAATGAGCTGGCACG 61170

Qy   2765 gcaacactattcactgatcaggaactgtccgccgaagattttgacaga 2812
           |||||||||||||||||||||||||||||||||||||||||||||
Db  61169 GCAACACTATTCACTGATCAGGAACTGTCCGCCGAAGATTTTGACAGA 61122
```

```
RESULT   2
CEY66A7/c
LOCUS       CEY66A7    197735 bp    DNA              HTG       03-DEC-1998
DEFINITION  Caenorhabditis elegans chromosome III clone Y66A7, *** SEQUENCING
            IN PROGRESS ***, in unordered pieces.
ACCESSION   AL022282
VERSION     AL022282.1  GI:3451532
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 197735)
  AUTHORS   Matthews,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1998) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Aug 25, 1998 this sequence version replaced gi:3250755.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .197735
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="III"
                     /clone="Y66A7"
BASE COUNT    62861 a  35824 c  35278 g  62783 t    989 others
ORIGIN
```

```
  Query Match           29.5%;  Score 1008.4;  DB 31;  Length 197735;
  Best Local Similarity   76.6%;  Pred. No. 1.9e-280;
  Matches 1508;  Conservative    0;  Mismatches    1;  Indels  459;  Gaps    4;

Qy   1304 attaccgtacccggaaccaattcaatacactgaatactctgcccgatcacccgtcgaatc 1363
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4043 ATTACAGTACCCGGAACCAATTCAATACACTGAATACTCTGCCCGATCACCCGTCGAATC 3984

Qy   1364 atctccttctccacctagaaatgagagaaaacgatcacgaatgagcacaacgactcctat 1423
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3983 ATCTCCTTCTCCACCTAGAAATGAGAGAAAACGATCACGAATGAGCACAACGACTCCTAT 3924

Qy   1424 gaagaatggaactagagatggatcttcgaaagttaccctttttgcgatgctcctagcagt 1483
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3923 GAAGAATGGAACTAGAGATGGATCTTCGAAAGTTACCCTTTTTGCGATGCTCCTAGCAGT 3864

Qy   1484 tctgatttttaatccgattg---------------------------------------- 1503
           ||||||||||||||||||
Db   3863 TCTGATTTTTAATCCGATTGGGTAAGCTTCTATTCATCTGAAAAATAGCGTTAAATCCTT 3804

Qy   1504 -----gattgctcgctggaagtgcgatattctcaaaagccgctgcagaagctccgattgc 1558
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3803 GTTTAGATTGCTCGCTGGAAGTGCGATATTCTCAAAAGCCGCTGCAGAAGCTCCGATTGC 3744

Qy   1559 ctccccgttcgagcatggaagagtgattgatgacccggatggaactagcactcggacgct 1618
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3743 CTCCCCGTTCGAGCATGGAAGAGTGATTGATGACCCGGATGGAACTAGCACTCGGACGCT 3684

Qy   1619 tttctgggaagggagtatcatcaatatgagctatgtctgggtgttcaacatcttaatgat 1678
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3683 TTTCTGGGAAGGGAGTATCATCAATATGAGCTATGTCTGGGTGTTCAACATCTTAATGAT 3624
```

```
Qy   1679 catatatgtggttgtcaaactgctgatccatggtgaccctgttcaagacttcatgtccgt 1738
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3623 CATATATGTGGTTGTCAAACTGCTGATCCATGGTGACCCTGTTCAAGACTTCATGTCCGT 3564

Qy   1739 ttcatggcagacttttgtgacgactcgagagaaggcgagagccgagttgaactctggaaa 1798
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3563 TTCATGGCAGACTTTTGTGACGACTCGAGAGAAGGCGAGAGCCGAGTTGAACTCTGGAAA 3504

Qy   1799 tttgaaagatgctcagagaaagttctgcgagtgtcttgcaacgttggatcgatcgcttcc 1858
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3503 TTTGAAAGATGCTCAGAGAAAGTTCTGCGAGTGTCTTGCAACGTTGGATCGATCGCTTCC 3444

Qy   1859 atcaccgggggttgattcggtgttttcggttggctgggaatgcgttcgacatcttttgaa 1918
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3443 ATCACCGGGGGTTGATTCGGTGTTTTCGGTTGGCTGGGAATGCGTTCGACATCTTTTGAA 3384

Qy   1919 ttggttgtggatcgggagatacatcgcaagaaggcgcaggtccaccacgaagcctgtctc 1978
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3383 TTGGTTGTGGATCGGGAGATACATCGCAAGAAGGCGCAGGTCCACCACGAAGCCTGTCTC 3324

Qy   1979 agtcgtttgtaggagtcatgcgcagactgcagttctctatcatgaaattcat-------- 2030
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3323 AGTCGTTTGTAGGAGTCATGCGCAGACTGCAGTTCTCTATCATGAAATTCATCAGGTTTA 3264

Qy   2031 ---------------------------------------------------------- 2030

Db   3263 CTGTCAAATTCCATAAATTTTGCATTCAAAAATATATTTACTAGAGTAATCCTACAGTAC 3204

Qy   2031 ---------------------------------------------------------- 2030

Db   3203 TGCTACAGTACCCCTACAGTACTACTACAGTACCCTGACCATATCCCCACTAACTCCAAA 3144

Qy   2031 ---------------------------------------------------------- 2030

Db   3143 CTAATATCCCTTCAAAAGACAAAAAGTCAATTTTTCCAAAACTACAGTAATCCTTCAGTA 3084

Qy   2031 ---------------------------------------------------------- 2030

Db   3083 CTCCTACGGTACCATTACAGTACTACTACAGTACCCTGACCATATCCCCCCACTAACTCC 3024

Qy   2031 ---------------------------------------------------------- 2030

Db   3023 GAACTAATATCCCTTCAAGAGACAGAAACATGTATTTTTTATGTATAAGTAATGATTTTT 2964

Qy   2031 -cagctccatctaatgggtatcactggaaacttcgaagacacctatgaaccatccgccct 2089
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2963 CCAGCTCCATCTAATGGGTATCACTGGAAACTTCGAAGACACCTATGAACCATCCGCCCT 2904

Qy   2090 aacgggcctcttcatgtccctctgtgcagtaaaccttgctgaagctgccggagcatcaaa 2149
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2903 AACGGGCCTCTTCATGTCCCTCTGTGCAGTAAACCTTGCTGAAGCTGCCGGAGCATCAAA 2844

Qy   2150 cgacggacttccacgcgccgtcatggctcagatctacatttctgcatccatccaatgccg 2209
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2843 CGACGGACTTCCACGCGCCGTCATGGCTCAGATCTACATTTCTGCATCCATCCAATGCCG 2784

Qy   2210 tttggctcttccgaacctactcgcaccattcttctcgggatacttttttacgaagagctcg 2269
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2783 TTTGGCTCTTCCGAACCTACTCGCACCATTCTTCTCGGGATACTTTTTTACGAAGAGCTCG 2724

Qy   2270 aaggcacgtgcgtcgagctccggagcactcggtgtcccatttgttatggatcttccatcc 2329
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2723 AAGGCACGTGCGTCGAGCTCCGGAGCACTCGGTGTCCCATTTGTTATGGATCTTCCATCC 2664

Qy   2330 agcgacaagaaagttcatgtcagatgcgaaaaggttggagcatgtgttgagctcgaagca 2389
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2663 AGCGACAAGAAAGTTCATGTCAGATGCGAAAAGGTTGGAGCATGTGTTGAGCTCGAAGCA 2604

Qy   2390 gaagcagttgagatttgggtcttttgtggaagatgagcaat------------------- 2430
          ||||||||||||||||||||||||||||||||||||||||
Db   2603 GAAGCAGTTGAGATTTGGGTCTTTTGTGGAAGATGAGCAATGTAAGTATTTTAAAATAAA 2544
```

```
Qy   2431 ----------------------------------------tatccccacttgctcgaatc 2450
                                                   ||||||||||||||||||||
Db   2543 ATAAAATAAAATAAAAAATAAATAAACTTTCATTTTCCAGTATCCCCACTTGCTCGAATC 2484

Qy   2451 cgaacaacgctgaaagtgtacctactctccaaacttgtacaggaacttgtcggtggtgac 2510
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2483 CGAACAACGCTGAAAGTGTACCTACTCTCCAAACTTGTACAGGAACTTGTCGGTGGTGAC 2424

Qy   2511 gagatctttacaaaaaatgtggaacgcatcctaaatgacaatgaccgtctcgatgatgaa 2570
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2423 GAGATCTTTACAAAAAATGTGGAACGCATCCTAAATGACAATGACCGTCTCGATGATGAA 2364

Qy   2571 gtagacgtggttgatgtttcaagacttttggtgacaatttcaacgcagtgcgctgccatt 2630
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2363 GTAGACGTGGTTGATGTTTCAAGACTTTTGGTGACAATTTCAACGCAGTGCGCTGCCATT 2304

Qy   2631 ttgactaatgagaaggatgagtcagcga------------------------------ 2658
          ||||||||||||||||||||||||||||
Db   2303 TTGACTAATGAGAAGGATGAGTCAGCGAGTTAGTGAACACGAATTTTAAGATTGAAAATG 2244

Qy   2659 --------------aattcggaacctggatctctcgaaacggagatgcttgttgcacatg 2704
                         ||||||||||||||||||||||||||||||||||||||||||||||
Db   2243 TTATAAATTTCTAGAATTCGGAACCTGGATCTCTCGAAACGGAGATGCTTGTTGCACATG 2184

Qy   2705 gtggacgcacgttctgacatgtggaatctattggaggagtaacaagaatgagctggcacg 2764
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2183 GTGGACGCACGTTCTGACATGTGGAATCTATTGGAGGAGTAACAAGAATGAGCTGGCACG 2124

Qy   2765 gcaacactattcactgatcaggaactgtccgccgaagattttgacaga 2812
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   2123 GCAACACTATTCACTGATCAGGAACTGTCCGCCGAAGATTTTGACAGA 2076
```

```
RESULT   3
CEY47D3/c
LOCUS       CEY47D3    337565 bp    DNA             HTG       02-SEP-1999
DEFINITION  Caenorhabditis elegans chromosome III clone Y47D3, *** SEQUENCING
            IN PROGRESS ***, in unordered pieces.
ACCESSION   Z98865
VERSION     Z98865.1  GI:4164282
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 337565)
  AUTHORS   Matthews,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1999) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Jan 19, 1999 this sequence version replaced gi:4056520.
            Order of segments is not known; 800 n's separate segments.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .337565
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="III"
                     /clone="Y47D3"
BASE COUNT   108087 a  59856 c  60930 g 107892 t    800 others
ORIGIN
```

```
  Query Match             29.5%;  Score 1008.4;  DB 31;  Length 337565;
  Best Local Similarity   76.6%;  Pred. No. 2.1e-280;
  Matches 1508;  Conservative    0;  Mismatches    1;  Indels  459;  Gaps    4;

Qy   1304 attaccgtacccggaaccaattcaatacactgaatactctgcccgatcacccgtcgaatc 1363
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 304686 ATTACAGTACCCGGAACCAATTCAATACACTGAATACTCTGCCCGATCACCCGTCGAATC 304627

Qy   1364 atctccttctccacctagaaatgagagaaaacgatcacgaatgagcacaacgactcctat 1423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 304626 ATCTCCTTCTCCACCTAGAAATGAGAGAAAACGATCACGAATGAGCACAACGACTCCTAT 304567

Qy   1424 gaagaatggaactagagatggatcttcgaaagttacccttttgcgatgctcctagcagt 1483
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 304566 GAAGAATGGAACTAGAGATGGATCTTCGAAAGTTACCCTTTTTGCGATGCTCCTAGCAGT 304507

Qy   1484 tctgatttttaatccgattg---------------------------------------- 1503
          ||||||||||||||||||||
Db 304506 TCTGATTTTTAATCCGATTGGGTAAGCTTCTATTCATCTGAAAAATAGCGTTAAATCCTT 304447

Qy   1504 -----gattgctcgctggaagtgcgatattctcaaaagccgctgcagaagctccgattgc 1558
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 304446 GTTTAGATTGCTCGCTGGAAGTGCGATATTCTCAAAAGCCGCTGCAGAAGCTCCGATTGC 304387

Qy   1559 ctccccgttcgagcatggaagagtgattgatgacccggatggaactagcactcggacgct 1618
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 304386 CTCCCCGTTCGAGCATGGAAGAGTGATTGATGACCCGGATGGAACTAGCACTCGGACGCT 304327

Qy   1619 tttctgggaagggagtatcatcaatatgagctatgtctgggtgttcaacatcttaatgat 1678
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 304326 TTTCTGGGAAGGGAGTATCATCAATATGAGCTATGTCTGGGTGTTCAACATCTTAATGAT 304267

Qy   1679 catatatgtggttgtcaaactgctgatccatggtgaccctgttcaagacttcatgtccgt 1738
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 304266 CATATATGTGGTTGTCAAACTGCTGATCCATGGTGACCCTGTTCAAGACTTCATGTCCGT 304207

Qy   1739 ttcatggcagacttttgtgacgactcgagagaaggcgagagccgagttgaactctggaaa 1798
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 304206 TTCATGGCAGACTTTTGTGACGACTCGAGAGAAGGCGAGAGCCGAGTTGAACTCTGGAAA 304147

Qy   1799 tttgaaagatgctcagagaaagttctgcgagtgtcttgcaacgttggatcgatcgcttcc 1858
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 304146 TTTGAAAGATGCTCAGAGAAAGTTCTGCGAGTGTCTTGCAACGTTGGATCGATCGCTTCC 304087

Qy   1859 atcaccgggggttgattcggtgttttcggttggctgggaatgcgttcgacatcttttgaa 1918
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 304086 ATCACCGGGGGTTGATTCGGTGTTTTCGGTTGGCTGGGAATGCGTTCGACATCTTTTGAA 304027

Qy   1919 ttggttgtggatcgggagatacatcgcaagaaggcgcaggtccaccacgaagcctgtctc 1978
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 304026 TTGGTTGTGGATCGGGAGATACATCGCAAGAAGGCGCAGGTCCACCACGAAGCCTGTCTC 303967

Qy   1979 agtcgtttgtaggagtcatgcgcagactgcagttctctatcatgaaattcat-------- 2030
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 303966 AGTCGTTTGTAGGAGTCATGCGCAGACTGCAGTTCTCTATCATGAAATTCATCAGGTTTA 303907

Qy   2031 ---------------------------------------------------------- 2030

Db 303906 CTGTCAAATTCCATAAATTTTGCATTCAAAAATATATTTACTAGAGTAATCCTACAGTAC 303847

Qy   2031 ---------------------------------------------------------- 2030

Db 303846 TGCTACAGTACCCCTACAGTACTACTACAGTACCCTGACCATATCCCCACTAACTCCAAA 303787

Qy   2031 ---------------------------------------------------------- 2030

Db 303786 CTAATATCCCTTCAAAAGACAAAAAGTCAATTTTTCCAAAACTACAGTAATCCTTCAGTA 303727

Qy   2031 ---------------------------------------------------------- 2030

Db 303726 CTCCTACGGTACCATTACAGTACTACTACAGTACCCTGACCATATCCCCCCACTAACTCC 303667
```

```
Qy   2031 ---------------------------------------------------------- 2030

Db 303666 GAACTAATATCCCTTCAAGAGACAGAAACATGTATTTTTTATGTATAAGTAATGATTTTT 303607

Qy   2031 -cagctccatctaatgggtatcactggaaacttcgaagacacctatgaaccatccgccct 2089
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 303606 CCAGCTCCATCTAATGGGTATCACTGGAAACTTCGAAGACACCTATGAACCATCCGCCCT 303547

Qy   2090 aacgggcctcttcatgtccctctgtgcagtaaaccttgctgaagctgccggagcatcaaa 2149
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 303546 AACGGGCCTCTTCATGTCCCTCTGTGCAGTAAACCTTGCTGAAGCTGCCGGAGCATCAAA 303487

Qy   2150 cgacggacttccacgcgccgtcatggctcagatctacatttctgcatccatccaatgccg 2209
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 303486 CGACGGACTTCCACGCGCCGTCATGGCTCAGATCTACATTTCTGCATCCATCCAATGCCG 303427

Qy   2210 tttggctcttccgaacctactcgcaccattcttctcgggatacttttacgaagagctcg 2269
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 303426 TTTGGCTCTTCCGAACCTACTCGCACCATTCTTCTCGGGATACTTTTTACGAAGAGCTCG 303367

Qy   2270 aaggcacgtgcgtcgagctccggagcactcggtgtcccatttgttatggatcttccatcc 2329
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 303366 AAGGCACGTGCGTCGAGCTCCGGAGCACTCGGTGTCCCATTTGTTATGGATCTTCCATCC 303307

Qy   2330 agcgacaagaaagttcatgtcagatgcgaaaaggttggagcatgtgttgagctcgaagca 2389
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 303306 AGCGACAAGAAAGTTCATGTCAGATGCGAAAAGGTTGGAGCATGTGTTGAGCTCGAAGCA 303247

Qy   2390 gaagcagttgagatttgggtcttttgtggaagatgagcaat------------------- 2430
          ||||||||||||||||||||||||||||||||||||||||
Db 303246 GAAGCAGTTGAGATTTGGGTCTTTTGTGGAAGATGAGCAATGTAAGTATTTTAAAATAAA 303187

Qy   2431 ----------------------------------------tatccccacttgctcgaatc 2450
                                                   ||||||||||||||||||||
Db 303186 ATAAAATAAAATAAAAAATAAATAAACTTTCATTTTCCAGTATCCCCACTTGCTCGAATC 303127

Qy   2451 cgaacaacgctgaaagtgtacctactctccaaacttgtacaggaacttgtcggtggtgac 2510
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 303126 CGAACAACGCTGAAAGTGTACCTACTCTCCAAACTTGTACAGGAACTTGTCGGTGGTGAC 303067

Qy   2511 gagatctttacaaaaaaatgtggaacgcatcctaaatgacaatgaccgtctcgatgatgaa 2570
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 303066 GAGATCTTTACAAAAAATGTGGAACGCATCCTAAATGACAATGACCGTCTCGATGATGAA 303007

Qy   2571 gtagacgtggttgatgtttcaagacttttggtgacaatttcaacgcagtgcgctgccatt 2630
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 303006 GTAGACGTGGTTGATGTTTCAAGACTTTTGGTGACAATTTCAACGCAGTGCGCTGCCATT 302947

Qy   2631 ttgactaatgagaaggatgagtcagcga---------------------------------- 2658
          ||||||||||||||||||||||||||||
Db 302946 TTGACTAATGAGAAGGATGAGTCAGCGAGTTAGTGAACACGAATTTTAAGATTGAAAATG 302887

Qy   2659 --------------aattcggaacctggatctctcgaaacggagatgcttgttgcacatg 2704
                         ||||||||||||||||||||||||||||||||||||||||||||
Db 302886 TTATAAATTTCTAGAATTCGGAACCTGGATCTCTCGAAACGGAGATGCTTGTTGCACATG 302827

Qy   2705 gtggacgcacgttctgacatgtggaatctattggaggagtaacaagaatgagctggcacg 2764
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 302826 GTGGACGCACGTTCTGACATGTGGAATCTATTGGAGGAGTAACAAGAATGAGCTGGCACG 302767

Qy   2765 gcaacactattcactgatcaggaactgtccgccgaagattttgacaga 2812
          ||||||||||||||||||||||||||||||||||||||||||||||
Db 302766 GCAACACTATTCACTGATCAGGAACTGTCCGCCGAAGATTTTGACAGA 302719
```

```
RESULT   4
CEH10N23/c
LOCUS       CEH10N23    36780 bp    DNA             HTG       22-SEP-1998
DEFINITION  Caenorhabditis elegans chromosome III clone H10N23, *** SEQUENCING
            IN PROGRESS ***, in unordered pieces.
ACCESSION   Z94157
VERSION     Z94157.1  GI:3377969
KEYWORDS    HTG; HTGS_PHASE1.
```

```
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 36780)
  AUTHORS   Sulston,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-1998) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Aug 3, 1998 this sequence version replaced gi:1945146.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .36780
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="III"
                     /clone="H10N23"
BASE COUNT     8697 a   5380 c   5339 g   8564 t   8800 others
ORIGIN
```

```
  Query Match             17.7%;  Score 604.8;  DB 31;  Length 36780;
  Best Local Similarity   93.4%;  Pred. No. 1.4e-163;
  Matches  661;  Conservative    0;  Mismatches    2;  Indels   45;  Gaps    1;

Qy    1304 attaccgtacccggaaccaattcaatacactgaatactctgcccgatcacccgtcgaatc 1363
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     708 ATTACAGTACCCGGAACCAATTCAATACACTGAATACTCTGCCCGATCACCCGTCGAATC 649

Qy    1364 atctccttctccacctagaaatgagagaaaacgatcacgaatgagcacaacgactcctat 1423
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     648 ATCTCCTTCTCCACCTAGAAATGAGAGAAAACGATCACGAATGAGCACAACGACTCCTAT 589

Qy    1424 gaagaatggaactagagatggatcttcgaaagttaccctttttgcgatgctcctagcagt 1483
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     588 GAAGAATGGAACTAGAGATGGATCTTCGAAAGTTACCCTTTTTGCGATGCTCCTAGCAGT 529

Qy    1484 tctgatttttaatccgattg---------------------------------------- 1503
           ||||||||||||||||||||
Db     528 TCTGATTTTTAATCCGATTGGGTAAGCTTCTATTCATCTGAAAAATAGCGTTAAATCCTT 469

Qy    1504 -----gattgctcgctggaagtgcgatattctcaaaagccgctgcagaagctccgattgc 1558
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     468 GTTTAGATTGCTCGCTGGAAGTGCGATATTCTCAAAAGCCGCTGCAGAAGCTCCGATTGC 409

Qy    1559 ctccccgttcgagcatggaagagtgattgatgacccggatggaactagcactcggacgct 1618
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     408 CTCCCCGTTCGAGCATGGAAGAGTGATTGATGACCCGGATGGAACTAGCACTCGGACGCT 349

Qy    1619 tttctgggaagggagtatcatcaatatgagctatgtctgggtgttcaacatcttaatgat 1678
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     348 TTTCTGGGAAGGGAGTATCATCAATATGAGCTATGTCTGGGTGTTCAACATCTTAATGAT 289

Qy    1679 catatatgtggttgtcaaactgctgatccatggtgaccctgttcaagacttcatgtccgt 1738
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     288 CATATATGTGGTTGTCAAACTGCTGATCCATGGTGACCCTGTTCAAGACTTCATGTCCGT 229

Qy    1739 ttcatggcagacttttgtgacgactcgagagaaggcgagagccgagttgaactctggaaa 1798
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db     228 TTCATGGCAGACTTTTGTGACGACTCGAGAGAAGGCGAGAGCCGAGTTGAATTCTGGAAA 169
```

```
Qy    1799 tttgaaagatgctcagagaaagttctgcgagtgtcttgcaacgttggatcgatcgcttcc 1858
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     168 TTTGAAAGATGCTCAGAGAAAGTTCTGCGAGTGTCTTGCAACGTTGGATCGATCGCTTCC 109

Qy    1859 atcaccgggggttgattcggtgttttcggttggctgggaatgcgttcgacatcttttgaa 1918
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     108 ATCACCGGGGGTTGATTCGGTGTTTTCGGTTGGCTGGGAATGCGTTCGACATCTTTTGAA 49

Qy    1919 ttggttgtggatcgggagatacatcgcaagaaggcgcaggtccaccac 1966
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db      48 TTGGTTGTGGATCGGGAGATACATCGCAAGAAGGCGCAGGTCCACCAC 1
```

```
RESULT   5
CEY48A6
LOCUS       CEY48A6    296699 bp    DNA             HTG       31-JUL-1998
DEFINITION  Caenorhabditis elegans chromosome III clone Y48A6, *** SEQUENCING
            IN PROGRESS ***, in unordered pieces.
ACCESSION   Z92854
VERSION     Z92854.1  GI:3218061
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 296699)
  AUTHORS   Gardner,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-1998) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Jun 13, 1998 this sequence version replaced gi:3204147.
            Order of segments is not known; 800 n's separate segments.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .296699
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="III"
                     /clone="Y48A6"
BASE COUNT    93035 a  52118 c  54334 g  90807 t   6405 others
ORIGIN
```

```
  Query Match             15.6%;  Score 535;  DB 31;  Length 296699;
  Best Local Similarity   90.7%;  Pred. No. 3.7e-143;
  Matches  614;  Conservative    1;  Mismatches    6;  Indels   56;  Gaps    2;

Qy    2799 aagattttgacagacaatctgggtttggcggttggccacgcgttgtgtgctcgcaagatt 2858
           || |||||  |   ||||||||||||||||||||||||||||||||||||||||||||||
Db   19474 AACATTTTTTCCAGCAATCTGGGTTTGGCGGTTGGCCACGCGTTGTGTGCTCGCAAGATT 19533

Qy    2859 tgcatagatgaccgagattccccgaaagtcagtcaatacgtgtgcattcacacaaagaag 2918
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19534 TGCATAGATGACCGAGATTCCCCGAAAGTCAGTCAATACGTGTGCATTCACACAAAGAAG 19593

Qy    2919 tcgctcgaatccctccgactattctccacatcatcgcgagcatcaggtgtggtgtctgga 2978
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19594 TCGCTCGAATCCCTCCGACTATTCTCCACATCATCGCGAGCATCAGGTGTGGTGTCTGGA 19653

Qy    2979 attcaggaaggtacacgccgaatggcctacgaatggattatgaactcgctgctcgacgcg 3038
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19654 ATTCAGGAAGGTACACGCCGAATGGCCTACGAATGGATTATGAACTCGCTGCTCGACGCG 19713
```

```
Qy   3039 tggcgttccaatctattcgcatcgaaaccctactggacacaaagcttcaagggacaatcc 3098
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19714 TGGCGTTCCAATCTATTCGCATCGAAACCCTACTGGACACAAAGCTTCAAGGGACAATCC 19773

Qy   3099 acgtttagtacgctttatcaagaggcgtataatcattatgcgattattaatgggacaagg 3158
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19774 ACGTTTAGTACGCTTTATCAAGAGGCGTATAATCATTATGCGATTATTAATGGGACAAGG 19833

Qy   3159 ggagattgttggaga---------------------------------------------- 3173
          |||||||||||||||
Db  19834 GGAGATTGTTGGAGAGTGAGTTATTACATTACTCCTCCAGTAACCTAAACTAAACTCTTT 19893

Qy   3174 ----------ctatttgtctacgagctcacgtgccgaatgctcaacggagccaacccaca 3223
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19894 ATCGTTTTAGCTATTTGTCTACGAGCTCACGTGCCGAATGCTCAACGGAGCCAACCCACA 19953

Qy   3224 agccacgtggtcaggcgyccgacgcgttcgatctacaaaaatggacgcggtccgaggaag 3283
          ||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db  19954 AGCCACGTGGTCAGGCGTCCGACGCGTTCGATCTACAAAAATGGACGCGGTCCGAGGAAG 20013

Qy   3284 agtgagcatgc-gacgctcggctcaaccggacgcatttcatcttcatacactggttaaac 3342
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db  20014 AGTGAGCATGCGGACGCTCGGCTCAACCGGACGCATTTCATCTTCATACACTGGTTAAAC 20073

Qy   3343 tacatacttctatggatctttgaattgaacaaaaaatgattttattcagaataatgataa 3402
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20074 TACATACTTCTATGGATCTTTGAATTGAACAAAAAATGATTTTATTCAGAATAATGATAA 20133

Qy   3403 atacgattatatataaa 3419
          |||||||||||||||||
Db  20134 ATACGATTATATATAAA 20150


RESULT   6
CEH10N23
LOCUS       CEH10N23    36780 bp    DNA             HTG       22-SEP-1998
DEFINITION  Caenorhabditis elegans chromosome III clone H10N23, *** SEQUENCING
            IN PROGRESS ***, in unordered pieces.
ACCESSION   Z94157
VERSION     Z94157.1  GI:3377969
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 36780)
  AUTHORS   Sulston,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-1998) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Aug 3, 1998 this sequence version replaced gi:1945146.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .36780
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="III"
                     /clone="H10N23"
BASE COUNT     8697 a   5380 c   5339 g   8564 t   8800 others
ORIGIN
```

```
  Query Match             12.4%;  Score 424.2;  DB 31;  Length 36780;
  Best Local Similarity   86.8%;  Pred. No. 3.2e-111;
  Matches  501;  Conservative    0;  Mismatches   23;  Indels   53;  Gaps    1;

Qy    544 agaatttcactcatgcaatgctatcttcaccacaccatacctcaatgacttctcaaccat 603
          ||||||||||||||||||||||||| |||||||||||||||||||| |||||||||
Db   5925 AGAATTTCACTCATGCAATGCTATCTCCACCACACCATACCTCAATGACTCCTCAACCAT 5984

Qy    604 atacagaagccatgggacatatcaacgggtacatgtctccatacgacca---------- 652
          |||||||||| |||| ||||||||||||||||||||||||||||||||
Db   5985 ATACAGAAGCAATGGAACATATCAACGGGTACATGTCTCCATACGACCAAGGTTTTCTTT 6044

Qy    653 ----------------------------------------agctcaaggcccatcagg 670
                                                   ||||||||||||||||||
Db   6045 TCCCCTTTACTGATAACCTTATATTGATCCTTAATAATTTCCAGCTCAAGGCCCATCAGG 6104

Qy    671 accatcatattactcacaacaccatcaatctccaccacctcatcaccaccatcaccaccc 730
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6105 ACCATCATATTACTCACAACACCATCAATCTCCACCACCTCATCACCACCATCACCACCC 6164

Qy    731 gatgccaaaaatccatgagaaccctgaacaagtggcatctccatcgattgaagatgctcc 790
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6165 GATGCCAAAAATCCATGAGAACCCTGAACAAGTGGCATCTCCATCGATTGAAGATGCTCC 6224

Qy    791 agagacgaaaccaactcatttggttgaaccacaaagtccaaaaagcccgcagaatatgaa 850
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6225 AGAGACGAAACCAACTCATTTGGTTGAACCACAAAGTCCAAAAAGCCCGCAGAATATGAA 6284

Qy    851 agaggagcttcttcggttactagttaacatgtctccgagtgaagttgaacggttaaagaa 910
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6285 AGAGGAGCTTCTTCGGTTACTAGTTAACATGTCTCCGAGTGAAGTTGAACGGTTAAAGAA 6344

Qy    911 taaaaaatcaggagcatgttcagcgacgaatgggccatcgaggagtaaggagaaggcggc 970
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6345 TAAAAAATCAGGAGCATGTTCAGCGACGAATGGGCCATCGAGGAGTAAGGAGAAGGCGGC 6404

Qy    971 gaagattgtgattcaggagacagcggaaggggatgaagatgaggatgatgaggatagtga 1030
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   6405 GAAGATTGTGATTCAGGAGACAGCCGAAGGGGATGAAGATGAGGATGATGAGGATAGTGA 6464

Qy   1031 ttccggggagactatgtctcagggaactactattatt 1067
          || ||   | |||||  ||  |         ||||||
Db   6465 TTACGCACACACTATCACTGCGAACTGACTCATTATT 6501


RESULT   7
I66494/c
LOCUS       I66494       7218 bp    DNA             PAT       28-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN


  Query Match             2.1%;  Score 71.8;  DB 5;  Length 7218;
  Best Local Similarity   11.7%;  Pred. No. 3.4e-09;
  Matches   58;  Conservative  231;  Mismatches  208;  Indels    0;  Gaps    0;

Qy    771 ccatcgattgaagatgctccagagacgaaaccaactcatttggttgaaccacaaagtcca 830
```

```
          | ||| ||  |||  | |  |||||   ||  |  |  |  ::  :::  : ::::   :
Db   1475 CTATCTATGCAAGTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRR 1416

Qy    831 aaaagcccgcagaatatgaaagaggagcttcttcggttactagttaacatgtctccgagt 890
          :::::   : :::: : ::::::::::       ::  :  ::  :: : :     :::
Db   1415 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1356

Qy    891 gaagttgaacggttaaagaataaaaaatcaggagcatgttcagcgacgaatgggccatcg 950
          ::::  ::: ::  :::::: ::::::  ::::: : :   :: :: ::: :::  :  :
Db   1355 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1296

Qy    951 aggagtaaggagaaggcggcgaagattgtgattcaggagacagcggaaggggatgaagat 1010
          ::::: :::::::::: :: :::::  : ::   :::::: :: :::::::::: :::::
Db   1295 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1236

Qy   1011 gaggatgatgaggatagtgattccggggagactatgtctcagggaactactattattgtt 1070
          ::::: :: ::::: :: ::    :::::::  : :    ::::::  :  :  :  :
Db   1235 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1176

Qy   1071 cgaagaccaaaaaccgagcgtcgtacggcacacaatctcatcgaaaagaagtatagatgc 1130
           :::::  :::::  ::: :  : : :: : : ::    :  ::::::::: : ::: :
Db   1175 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1116

Qy   1131 tcaataaatgatcgaattcaacagctgaaagtacttttgtgtggggatgaagctaagctt 1190
            :: ::: ::  :::   :: ::  ::::: :     : : ::::: :|    |||||
Db   1115 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATCGCAAGCTC 1056

Qy   1191 tcaaaatcggcaacactacgacgggctattgaacatatcgaggaggttgaacacgagaat 1250
           |   | | ||| |    |   |   || |      | ||    |       | || ||
Db   1055 CCTCGACCTGCAGCCAAGCTCGGAATTAATTCTGTGAGCGTATGGCAAACGAAGGAAAAA 996

Qy   1251 caggtgttgaagcatca 1267
           || | | | |||  ||
Db    995 TAGTTATAGTAGCCGCA 979


RESULT   8
CGU09103
LOCUS       CGU09103     3816 bp    mRNA            ROD       26-JUL-1994
DEFINITION  Cricetulus griseus sterol regulatory element binding protein-1
            (SREBP-1) mRNA, complete cds.
ACCESSION   U09103
VERSION     U09103.1  GI:516002
KEYWORDS    .
SOURCE      Chinese hamster.
  ORGANISM  Cricetulus griseus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
REFERENCE   1  (bases 1 to 3816)
  AUTHORS   Sato,R., Yang,J., Wang,X., Evans,M.J., Ho,Y.K., Goldstein,J.L. and
            Brown,M.S.
  TITLE     Assignment of the membrane attachment, DNA binding, and
            transcriptional activation domains of sterol regulatory binding
            protein-1 (SREBP-1)
  JOURNAL   J. Biol. Chem. 269, 17267-17273 (1994)
  MEDLINE   94274723
REFERENCE   2  (bases 1 to 3816)
  AUTHORS   Yang,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-1994) Jianxin Yang, Molecular Genetics,
            University of Texas Southwestern Medical Center, 5323 Harry Hines
            Blvd., Dallas, TX 75237-9046, USA
FEATURES             Location/Qualifiers
     source          1. .3816
                     /organism="Cricetulus griseus"
                     /db_xref="taxon:10029"
                     /cell_line="CHO"
                     /cell_type="CHO K1"
     CDS             31. .3432
                     /note="SREBP-1"
                     /codon_start=1
                     /product="sterol regulatory element binding protein-1"
                     /protein_id="AAA20085.1"
                     /db_xref="GI:516003"
                     /translation="MDELPFGEAAVEQALDELGELDAALLTDIQDMLQLINNQDSDFP
                     GLFDSPYAGGGAGDTEPTSPGANSPESLSSPASLGSSLEAFLGEPKATPASLSPVPSA
                     STALKMYPSVPPFSPGPGIKEEPVPLTILQPPAAQPSPGTLLPPSFPPPPLQLSPAPV
                     LGYSSLPSGFSGTLPGNTQQPPSSLSLASAPGVSPISLHTQVQSSASQQPLPASTAPR
                     TTTVTSQIQRVPVVLQPHFIKADSLLLTTVKTDTGATMKTAGISTLAPGTAVQAGPLQ
                     TLVSGGTILATVPLVVDTDKLPIHRLAAGSKALGSAQSRGEKRTAHNAIEKRYRSSIN
                     DKIVELKDLVVGTEAKLNKSAVLRKAIDYIRFLQHSNQKLKQENLALRNAAHKSKSLK
                     DLVSACGSAGGTDVAMEGVKPEVVDTLTPPPSDAGSPSQSSPLSLGSRGSSSGGSDSE
                     PDSPVFEDSQVKAQRLHSHGMLDRSRLALCALVFLCLTCNPLASLFGWGIPGPSSASG
                     AHHSSGRSMLEAESRDGSNWTQWLLPPLVWLANGLLVLACLALLFVYGEPVTRPHTSP
                     AVHFWRHRKQADLDLARGDFAQAAOQLWLALQALGRPLPTSNLDLACSLLWNLIRHLL
                     QRLWVGRWLAGRAGGLRRDCGLRMDARASARDAALVYHKLHQLHAMGKYTGGHLIASN
                     LALSALNLAECAGDAVSMATLAEIYVAAALRVKTSLPRALHFLTRFFLSSARQACLAQ
                     SGSVPLAMQWLCHPVGHRFFVDGDWAVHGAPQESLYSVAGNPVDPLAQVTRLFCEHLL
                     ERALNCIAQPSPGTADGDREFSDALGYLQLLNRCSDAVGTPACSFSVSSSMASTTGTD
                     PVAKWWASLTAVVIHWLRRDEEAAERLYPLVERMPHVLQETERPLPKAALYSFKAARA
                     LLDHRKVESGPASLAICEKASGYLRDSLAAPPTGSSIDKAMQLLLCDLLLVARTSMWQ
                     RQQSPASAQVAHSASNGSQASALELRGFQQDLSSLRRLAQNFRPAMRRVFLHEATARL
                     MAGASPARTHQLLDRSLRRRAGSSGKGGTVAELEPRPTWREHTEALLLASCYLPPAFL
                     SAPGQQMSMLAEAARTVEKLGDHRLLLDCQQMLLRLGGGTTVTSS"
     polyA_signal    3793. .3799
     polyA_site      3816
                     /note="13 A residues"
BASE COUNT      770 a   1188 c   1061 g    797 t
ORIGIN


  Query Match             1.7%;  Score 57.4;  DB 12;  Length 3816;
  Best Local Similarity   53.3%;  Pred. No. 4.5e-05;
  Matches  121;  Conservative    0;  Mismatches  106;  Indels    0;  Gaps    0;

Qy   1086 gagcgtcgtacggcacacaatctcatcgaaaagaagtatagatgctcaataaatgatcga 1145
           |||   || || || ||||||  ||| || |||   ||  | | ||| || |||||
Db    979 GAGAAGCGCACAGCCCACAATGCCATTGAGAAGCGCTACCGTTCCTCTATCAATGACAAG 1038

Qy   1146 attcaacagctgaaagtacttttgtgtggggatgaagctaagctttcaaaatcggcaaca 1205
           |||    |||| |||||  ||  ||   ||   ||| || |||||    ||||| ||
Db   1039 ATTGTGGAGCTCAAAGACCTGGTGGTGGGCACTGAGGCAAAGCTGAATAAATCTGCCGTC 1098

Qy   1206 ctacgacgggctattgaacatatcgaggaggttgaacacgagaatcaggtgttgaagcat 1265
           | ||   ||| || ||  |||||       |  | |||   || |||  | | |||||
Db   1099 TTGCGCAAGGCCATCGACTATATCCGCTTCTTACAGCACAGCAACCAGAAGCTCAAGCAG 1158

Qy   1266 catgttgaacaaatgagaaagacactgcagaataatcgattaccgta 1312
           |         || ||||  |   || || |   ||  | | |
Db   1159 GAGAACCTGGCCCTGCGAAATGCCGCTCACAAAAGCAAATCCCTGAA 1205


RESULT   9
AB017337
LOCUS       AB017337     1209 bp    mRNA            ROD       14-APR-1999
DEFINITION  Mus musculus mRNA for sterol regulatory element-binding protein-1
            (SREBP-1), partial cds.
ACCESSION   AB017337
VERSION     AB017337.1  GI:4240011
KEYWORDS    sterol regulatory element-binding protein-1 (SREBP-1).
SOURCE      Mus musculus cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Inoue,J. and Sato,R.
  TITLE     A novel splicing isoform of mouse sterol regulatory element-binding
            protein-1 (SREBP-1)
  JOURNAL   Biosci. Biotechnol. Biochem. 63 (1), 243-245 (1999)
  MEDLINE   99161303
REFERENCE   2  (bases 1 to 1209)
  AUTHORS   Inoue,J. and Sato,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-AUG-1998) to the DDBJ/EMBL/GenBank databases. Jun
            Inoue, Graduate School of Pharmaceutical Sciences, Osaka
            University, Department of Biochemistry and Molecular Biology;
```

```
          1-6,Yamada-oka,, Suita, Osaka 565-0871, Japan
          (E-mail:inoue@phs.osaka-u.ac.jp, Tel:06-879-8188)
FEATURES             Location/Qualifiers
     source          1. .1209
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     gene            1. .1209
                     /gene="SREBF-1"
     CDS             <1. .>1209
                     /gene="SREBF-1"
                     /codon_start=1
                     /product="sterol regulatory element-binding protein-1
                     (SREBP-1)"
                     /protein_id="BAA74795.1"
                     /db_xref="GI:4240012"
                     /translation="DFPGLFDAPYAGGETGDTGPSSPGANSPESFSSASLASSLEAFL
                     GGPKVTPAPLSPPPSAPAALKMYPSVSPFSPGPGIKEEPVPLTILQPAAPQPSPGTLL
                     PPSFPAPPVQLSPAPVLGYSSLPSGFSGTLPGNTQQPPSSLPLAPAPGVLPTPALHTQ
                     VQSLASQQPLPASAAPRTNTVTSQVQQVPVVLQPHFIKADSLLLTAVKTDAGATVKTA
                     GISTLAPGTAVQAGPLQTLVSGGTILATVPLVVDTDKLPIHRLAAGSKALGSAQSRGE
                     KRTAHNAIEKRYRSSINDKIVELKDLVVGTEAKLNKSAVLRKAIDYIRFLQHSNQKLK
                     QENLTLRSAHKSKSLKDLVSACGSGGGTDVSMEGMKPEVVETLTPPPSDAGSPSQSSP
                     LSFGSRASSSG"
BASE COUNT      261 a    416 c    311 g    221 t
ORIGIN


  Query Match             1.7%;  Score 56.8;  DB 12;  Length 1209;
  Best Local Similarity   51.6%;  Pred. No. 5.3e-05;
  Matches  130;  Conservative    0;  Mismatches  122;  Indels    0;  Gaps    0;

Qy   1086 gagcgtcgtacggcacacaatctcatcgaaaagaagtatagatgctcaataaatgatcga 1145
          |||   || || || ||||||  ||| || |||   ||  | |  || || |||||
Db    826 GAGAAGCGCACAGCCCACAATGCCATTGAGAAGCGCTACCGGTCTTCTATCAATGACAAG 885

Qy   1146 attcaacagctgaaagtacttttgtgtggggatgaagctaagctttcaaaatcggcaaca 1205
          |||    |||| ||||  ||  ||   ||   |||||| |||||    ||||| ||
Db    886 ATTGTGGAGCTCAAAGACCTGGTGGTGGGCACTGAAGCAAAGCTGAATAAATCTGCTGTC 945

Qy   1206 ctacgacgggctattgaacatatcgaggaggttgaacacgagaatcaggtgttgaagcat 1265
          | ||   ||| || ||  | |||       |  | |||   || |||  | | |||||
Db    946 TTGCGCAAGGCCATCGACTACATCCGCTTCTTGCAGCACAGCAACCAGAAGCTCAAGCAG 1005

Qy   1266 catgttgaacaaatgagaaagacactgcagaataatcgattaccgtacccggaaccaatt 1325
          |          |  |||   |||   | |  ||   | |   | ||| ||   ||  |
Db   1006 GAGAACCTGACCCTACGAAGTGCACACAAAAGCAAATCACTGAAGGACCTGGTGTCAGCT 1065

Qy   1326 caatacactgaa 1337
               || || |
Db   1066 TGTGGCAGTGGA 1077


RESULT  10
CGU22819
LOCUS       CGU22819     1767 bp    mRNA            ROD       26-JAN-1996
DEFINITION  Cricetulus griseus SRD-3 mutant sterol regulatory element binding
            protein-2 (SREBP-2) mRNA, complete cds.
ACCESSION   U22819
VERSION     U22819.1  GI:841319
KEYWORDS    .
SOURCE      Chinese hamster.
  ORGANISM  Cricetulus griseus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
REFERENCE   1  (bases 1 to 1767)
  AUTHORS   Yang,J., Brown,M.S., Ho,Y.K. and Goldstein,J.L.
  TITLE     Three different rearrangements in a single intron truncate sterol
            regulatory element binding protein-2 and produce sterol-resistant
            phenotype in three cell lines. Role of introns in protein evolution
  JOURNAL   J. Biol. Chem. 270 (20), 12152-12161 (1995)
  MEDLINE   95263566
REFERENCE   2  (bases 1 to 1767)
  AUTHORS   Yang,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-1995) Jianxin Yang, Molecular Genetics,
            University of Texas Southwestern Medical Center, 5323 Harry Hine
            Blvd, Dallas, TX 75235, USA
FEATURES             Location/Qualifiers
     source          1. .1767
                     /organism="Cricetulus griseus"
                     /db_xref="taxon:10029"
                     /cell_line="SRD-3"
     gene            141. .1526
                     /gene="SREBP-2"
     CDS             141. .1526
                     /gene="SREBP-2"
                     /codon_start=1
                     /product="mutant sterol regulatory element binding
                     protein-2"
                     /protein_id="AAA85719.1"
                     /db_xref="GI:841320"
                     /translation="MDESSELGGLETMDTLTELGDELTLGDIDEMLQFVSNQVGEFPD
                     LFSEQLCSSFPGGGSSGSSSSSNSSSSSGSNSRGNGGAATDPGVQRSFSQVPLPTFSP
                     STASPQALALQVKVSPTPPRATPVLQPRPQPQPQPQPSAQLQQQTVMITPTFSTAPQT
                     RIIQQPLIYQNAATSFQVLQAQVQSLVTSSQVQPVTIQQQVQTVQAQRVLTQTANGTL
                     QTLAPATVQTVAAPQVQQVPVLVQPQIIKTDSLVLTTLKTDGSPVMAAVQNPALTALT
                     TPLQTGALQVPTLVGSNGTILTTMPVMMGQEKVPIKQVPGGVKQLEPPKEGERRTTHN
                     IIEKRYRSSINDKIIELKDLVMGTDAKMHKSGVLRKAIDYIKYLQQVNHKLRQENMVL
                     KLANQKNKLLKGIDLGSLVDSDVDLKIEDFNQNVLLMSPPASDSGSQAGFSPYSIDSE
                     PGSPLLDDAKV"
BASE COUNT      456 a    485 c    473 g    353 t
ORIGIN


  Query Match             1.6%;  Score 55.6;  DB 12;  Length 1767;
  Best Local Similarity   53.2%;  Pred. No. 0.00013;
  Matches  118;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy   1079 aaaaaccgagcgtcgtacggcacacaatctcatcgaaaagaagtatagatgctcaataaa 1138
          | ||   ||| |  | ||  |||||||  |||| ||||||   ||| ||| ||| || ||
Db   1115 AGAAGGAGAGAGGAGGACAACACACAACATCATTGAAAAGCGATATCGATCCTCTATCAA 1174

Qy   1139 tgatcgaattcaacagctgaaagtacttttgtgtggggatgaagctaagctttcaaaatc 1198
          |||   |||   | || |||| |   |  |    ||    || || ||| |    |||||
Db   1175 TGACAAAATCATAGAGTTGAAGGACTTGGTCATGGGAACAGATGCCAAGATGCACAAATC 1234

Qy   1199 ggcaacactacgacgggctattgaacatatcgaggaggttgaacacgagaatcaggtgtt 1258
           |      ||  |   ||| ||||| | ||| |  |  |  | || |  |||||   | |
Db   1235 TGGCGTTCTGAGGAAGGCCATTGATTACATCAAATATTTGCAGCAGGTCAATCACAAGCT 1294

Qy   1259 gaagcatcatgttgaacaaatgagaaagacactgcagaataa 1300
             ||  |         | |||    | ||   ||||| ||
Db   1295 TCGTCAGGAGAACATGGTACTGAAGCTGGCAAATCAGAAAAA 1336


RESULT  11
CGU12329
LOCUS       CGU12329     1914 bp    mRNA            ROD       10-AUG-1995
DEFINITION  Cricetulus griseus mutant sterol regulatory element binding
            protein-2 (mutant SREBP-2) mRNA, complete cds.
ACCESSION   U12329
VERSION     U12329.1  GI:551503
KEYWORDS    .
SOURCE      Chinese hamster.
  ORGANISM  Cricetulus griseus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
REFERENCE   1  (bases 1 to 1914)
  AUTHORS   Yang,J., Sato,R., Goldstein,J.L. and Brown,M.S.
  TITLE     Sterol-resistant transcription in CHO cells caused by gene
            rearrangement that truncates SREBP-2
  JOURNAL   Genes Dev. 8 (16), 1910-1919 (1994)
  MEDLINE   95047343
REFERENCE   2  (bases 1 to 1914)
  AUTHORS   Yang,J.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (15-JUL-1994) Jianxin Yang, Molecular Genetics,
            University of Texas, Southwestern Medical Center, 5323 Harry Hines
            Blvd., Dallas, TX 75235-9046, USA
FEATURES             Location/Qualifiers
     source          1. .1914
                     /organism="Cricetulus griseus"
                     /db_xref="taxon:10029"
                     /clone="pMThamSREBP-2"
                     /cell_line="SRD-1"
                     /cell_type="hybrid Chinese hamster ovarian cells"
                     /note="hybrid CHO cells include a tranlocation of the
                     SREBP-2 gene"
     gene            155. .1570
                     /gene="mutant SREBP-2"
     CDS             155. .1570
                     /gene="mutant SREBP-2"
                     /codon_start=1
                     /product="mutant sterol regulatory element binding
                     protein-2"
                     /protein_id="AAA74140.1"
                     /db_xref="GI:551504"
                     /translation="MDESSELGGLETMDTLTELGDELTLGDIDEMLQFVSNQVGEFPD
                     LFSEQLCSSFPGGGSSGSSSSSNSSSSSGSNSRGNGGAATDPGVQRSFSQVPLPTFSP
                     STASPQALALQVKVSPTPPRATPVLQPRPQPQPQPQPSAQLQQQTVMITPTFSTAPQT
                     RIIQQPLIYQNAATSFQVLQAQVQSLVTSSQVQPVTIQQQVQTVQAQRVLTQTANGTL
                     QTLAPATVQTVAAPQVQQVPVLVQPQIIKTDSLVLTTLKTDGSPVMAAVQNPALTALT
                     TPLQTGALQVPTLVGSNGTILTTMPVMMGQEKVPIKQVPGGVKQLEPPKEGERRTTHN
                     IIEKRYRSSINDKIIELKDLVMGTDAKMHKSGVLRKAIDYIKYLQQVNHKLRQENMVL
                     KLANQKNKLLKGIDLGSLVDSDVDLKIEDFNQNVLLMSPPASDSGSQAGFSPYSIDSE
                     PGSPLLDDAKMMKVLQVKSPR"
BASE COUNT      495 a    540 c    501 g    378 t
ORIGIN


  Query Match             1.6%;  Score 55.6;  DB 12;  Length 1914;
  Best Local Similarity   53.2%;  Pred. No. 0.00013;
  Matches  118;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy   1079 aaaaaccgagcgtcgtacggcacacaatctcatcgaaaagaagtatagatgctcaataaa 1138
          | ||   ||| |  | ||  |||||||  |||| ||||||   ||| ||| ||| || ||
Db   1129 AGAAGGAGAGAGGAGGACAACACACAACATCATTGAAAAGCGATATCGATCCTCTATCAA 1188

Qy   1139 tgatcgaattcaacagctgaaagtacttttgtgtggggatgaagctaagctttcaaaatc 1198
          |||   |||   | || |||| |   |  |    ||    || || ||| |    |||||
Db   1189 TGACAAAATCATAGAGTTGAAGGACTTGGTCATGGGAACAGATGCCAAGATGCACAAATC 1248

Qy   1199 ggcaacactacgacgggctattgaacatatcgaggaggttgaacacgagaatcaggtgtt 1258
           |     ||  |   ||| |||||  | ||| |  |  |  | || |  |||||   | |
Db   1249 TGGCGTTCTGAGGAAGGCCATTGATTACATCAAATATTTGCAGCAGGTCAATCACAAGCT 1308

Qy   1259 gaagcatcatgttgaacaaatgagaaagacactgcagaataa 1300
             ||  |          | |||    | ||   ||||| ||
Db   1309 TCGTCAGGAGAACATGGTACTGAAGCTGGCAAATCAGAAAAA 1350


RESULT  12
CGU22818
LOCUS       CGU22818     4425 bp    mRNA            ROD       26-JAN-1996
DEFINITION  Cricetulus griseus SRD-2 mutant sterol regulatory element binding
            protein-2 (SREBP-2) mRNA, complete cds.
ACCESSION   U22818
VERSION     U22818.1  GI:841317
KEYWORDS    .
SOURCE      Chinese hamster.
  ORGANISM  Cricetulus griseus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
REFERENCE   1  (bases 1 to 4425)
  AUTHORS   Yang,J., Brown,M.S., Ho,Y.K. and Goldstein,J.L.
  TITLE     Three different rearrangements in a single intron truncate sterol
            regulatory element binding protein-2 and produce sterol-resistant
            phenotype in three cell lines. Role of introns in protein evolution
  JOURNAL   J. Biol. Chem. 270 (20), 12152-12161 (1995)
  MEDLINE   95263566
REFERENCE   2  (bases 1 to 4425)
  AUTHORS   Yang,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-1995) Jianxin Yang, Molecular Genetics,
            University of Texas Southwestern Medical Center, 5323 Harry Hine
            Blvd, Dallas, TX 75235, USA
FEATURES             Location/Qualifiers
     source          1. .4425
                     /organism="Cricetulus griseus"
                     /db_xref="taxon:10029"
                     /cell_line="SRD-2"
     gene            138. .2657
                     /gene="SREBP-2"
     CDS             138. .2657
                     /gene="SREBP-2"
                     /codon_start=1
                     /product="mutant sterol regulatory element binding
                     protein-2"
                     /protein_id="AAA85718.1"
                     /db_xref="GI:841318"
                     /translation="MDESSELGGLETMDTLTELGDELTLGDIDEMLQFVSNQVGEFPD
                     LFSEQLCSSFPGGGSSGSSSSSNSSSSSGSNSRGNGGAATDPGVQRSFSQVPLPTFSP
                     STASPQALALQVKVSPTPPRATPVLQPRPQPQPQPQPSAQLQQQTVMITPTFSTAPQT
                     RIIQQPLIYQNAATSFQVLQAQVQSLVTSSQVQPVTIQQQVQTVQAQRVLTQTANGTL
                     QTLAPATVQTVAAPQVQQVPVLVQPQIIKTDSLVLTTLKTDGSPVMAAVQNPALTALT
                     TPLQTGALQVPTLVGSNGTILTTMPVMMGQEKVPIKQVPGGVKQLEPPKEGERRTTHN
                     IIEKRYRSSINDKIIELKDLVMGTDAKMHKSGVLRKAIDYIKYLQQVNHKLRQENMVL
                     KLANQKNKLLKGIDLGSLVDSDVDLKIEDFNQNVLLMSPPASDSGSQAGFSPYSIDSE
                     PGSPLLDDAKKFQAPAKEICVGCQKTVYPMERLLANQQVFHISCFRCSYCNNKLSLGT
                     YASLHGQIYCKPHFNQLFKSKGNYDEGFGHRPHKDLWANKSENEETSERSPQSPDVGE
                     IPHNSGVEDAPISKVGVLAANMEAKASPQREKEDKPAETKKLRIAWPPPAELGSSGST
                     MEEGFKGSKPKWPPEDETSKTEALGDVDLDLKKLRRSSSLKERSRPFTVAASFRTSSL
                     KSPKTLSPLRKGWSMSEQNEESGGGIVTAGKQTENAKAPGKKENVRKSRWQREDSARE
                     EVARRSKDIHGVELESENFIENGANIYEGDSGDLVQQSPLDPKSPSWSGFVDTTPTEE
                     FTTQNQKSHDVGFREGEVVRELSVEEQIKRNRYYDEDEDEE"
BASE COUNT     1236 a   1086 c   1101 g   1002 t
ORIGIN


  Query Match             1.6%;  Score 55.6;  DB 12;  Length 4425;
  Best Local Similarity   53.2%;  Pred. No. 0.00015;
  Matches  118;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy   1079 aaaaaccgagcgtcgtacggcacacaatctcatcgaaaagaagtatagatgctcaataaa 1138
          | ||   ||| |  | ||  |||||||  |||| ||||||   ||| ||| ||| || ||
Db   1112 AGAAGGAGAGAGGAGGACAACACACAACATCATTGAAAAGCGATATCGATCCTCTATCAA 1171

Qy   1139 tgatcgaattcaacagctgaaagtacttttgtgtggggatgaagctaagctttcaaaatc 1198
          |||   |||   | || |||| |   |  |    ||    || || ||| |    |||||
Db   1172 TGACAAAATCATAGAGTTGAAGGACTTGGTCATGGGAACAGATGCCAAGATGCACAAATC 1231

Qy   1199 ggcaacactacgacgggctattgaacatatcgaggaggttgaacacgagaatcaggtgtt 1258
           |     ||  |   ||| |||||  | ||| |  |  |  | || |  |||||   | |
Db   1232 TGGCGTTCTGAGGAAGGCCATTGATTACATCAAATATTTGCAGCAGGTCAATCACAAGCT 1291

Qy   1259 gaagcatcatgttgaacaaatgagaaagacactgcagaataa 1300
             ||  |          | |||    | ||   ||||| ||
Db   1292 TCGTCAGGAGAACATGGTACTGAAGCTGGCAAATCAGAAAAA 1333


RESULT  13
GU12330
LOCUS       GU12330      4987 bp    mRNA            ROD       10-AUG-1995
DEFINITION  Cricetulus griseus sterol regulatory element binding protein-2
            (SREBP-2) mRNA, complete cds.
ACCESSION   U12330
VERSION     U12330.1  GI:551505
KEYWORDS    .
SOURCE      Chinese hamster.
  ORGANISM  Cricetulus griseus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
```

```
REFERENCE   1  (bases 1 to 4987)
  AUTHORS   Yang,J., Sato,R., Goldstein,J.L. and Brown,M.S.
  TITLE     Sterol-resistant transcription in CHO cells caused by gene
            rearrangement that truncates SREBP-2
  JOURNAL   Genes Dev. 8 (16), 1910-1919 (1994)
  MEDLINE   95047343
REFERENCE   2  (bases 1 to 4987)
  AUTHORS   Yang,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-1994) Jianxin Yang, Molecular Genetics,
            University of Texas, Southwestern Medical Center, 5323 Harry Hines
            Blvd., Dallas, TX 75235-9046, USA
FEATURES             Location/Qualifiers
     source          1. .4987
                     /organism="Cricetulus griseus"
                     /db_xref="taxon:10029"
                     /clone="phamSREBP-2"
                     /cell_line="SRD-1 Cells"
                     /cell_type="Chinese hamster ovarian cells"
     gene            159. .3578
                     /gene="SREBP-2"
     CDS             159. .3578
                     /gene="SREBP-2"
                     /codon_start=1
                     /product="sterol regulatory element binding protein-2"
                     /protein_id="AAA74141.1"
                     /db_xref="GI:551506"
                     /translation="MDESSELGGLETMDTLTELGDELTLGDIDEMLQFVSNQVGEFPD
                     LFSEQLCSSFPGGGSSGSSSSSNSSSSSGSNSRGNGGAATDPGVQRSFSQVPLPTFSP
                     STASPQALALQVKVSPTPPRATPVLQPRPQPQPQPQPSAQLQQQTVMITPTFSTAPQT
                     RIIQQPLIYQNAATSFQVLQAQVQSLVTSSQVQPVTIQQQVQTVQAQRVLTQTANGTL
                     QTLAPATVQTVAAPQVQQVPVLVQPQIIKTDSLVLTTLKTDGSPVMAAVQNPALTALT
                     TPLQTGALQVPTLVGSNGTILTTMPVMMGQEKVPIKQVPGGVKQLEPPKEGERRTTHN
                     IIEKRYRSSINDKIIELKDLVMGTDAKMHKSGVLRKAIDYIKYLQQVNHKLRQENMVL
                     KLANQKNKLLKGIDLGSLVDSDVDLKIEDFNQNVLLMSPPASDSGSQAGFSPYSIDSE
                     PGSPLLDDAKVKDEPDSPPVALGMVDRSRILLCVLTFLGLSFSPLTSLLQWGGAHDTD
                     QHPYSGSGRSVLSLESGSGGWFDWMMPTLLLWLVNGVIVLSVFVKLLVHGEPVIRPHT
                     RSSVTFWRHRKQADLDLARGDFAAAAANLQTCLSVLGRALPTSRLDLACSLSWNVIRY
                     SLQKLRLVRWILKKVFQRWRATPATAAGFEDEAKSSARDAALAYHRLHQLHITGKLPA
                     GSTCSDVHMALCAVNLAECAEEKIPPSTLIEIHLTAAMGLKTGCGGKLGFLASYFLNR
                     AQSLCGPEHSTVPDSLRWLCHPLGQKFFMERSWSIKSAAKESLYCAQRNPADPIAQVH
                     QAFCKNLLERAVESLVKPQSKKKSGDQEDESCEFSSALEYLRLLHSFVDSVGFVTPPF
                     SSSSVLKSALGPDIICRWWTSAVTMAISWLQGDDAAVRSHFTEVERIPKALEVTESPL
                     VKAVFYACRAMHASLSGKGDGQQNSFCHCERASGHLWNSLNVSGATSDPSLNHVVQLL
                     TCDLLLSLRTALWQKQAGASQGLGETYHASGTELAGFQRDLGSLRRLAHSFRPAYRKV
                     FLHEPTVRLMAGANPTRTHQLLEHSLRRRTSQNTKHGEIDTWPGQRERATAILLACRH
                     LPLSFLSSPGQRAVLLAEAARTLEKVGDRRSCNDCQQMIVKLGGGTAIAAS"
BASE COUNT     1077 a   1394 c   1328 g   1188 t
ORIGIN
```

```
  Query Match             1.6%;  Score 55.6;  DB 12;  Length 4987;
  Best Local Similarity   53.2%;  Pred. No. 0.00016;
  Matches  118;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy   1079 aaaaaccgagcgtcgtacggcacacaatctcatcgaaaagaagtatagatgctcaataaa 1138
          | ||   ||| |  | ||  |||||||  |||| ||||||   ||| ||| ||| || ||
Db   1133 AGAAGGAGAGAGGAGGACAACACACAACATCATTGAAAAGCGATATCGATCCTCTATCAA 1192

Qy   1139 tgatcgaattcaacagctgaaagtacttttgtgtggggatgaagctaagctttcaaaatc 1198
          |||   |||   | || |||| |   |  |   ||    || || ||| |    |||||
Db   1193 TGACAAAATCATAGAGTTGAAGGACTTGGTCATGGGAACAGATGCCAAGATGCACAAATC 1252

Qy   1199 ggcaacactacgacgggctattgaacatatcgaggaggttgaacacgagaatcaggtgtt 1258
            |     ||  |  ||| |||||  | ||| |  |  |  | || |  |||||   | |
Db   1253 TGGCGTTCTGAGGAAGGCCATTGATTACATCAAATATTTGCAGCAGGTCAATCACAAGCT 1312

Qy   1259 gaagcatcatgttgaacaaatgagaaagacactgcagaataa 1300
             ||  |         | |||    | ||   ||||| ||
Db   1313 TCGTCAGGAGAACATGGTACTGAAGCTGGCAAATCAGAAAAA 1354
```

```
RESULT  14
```

```
RATADD1A
LOCUS       RATADD1A     3115 bp    mRNA            ROD       08-SEP-1993
DEFINITION  Rat add1 mRNA sequence.
ACCESSION   L16995
VERSION     L16995.1  GI:294509
KEYWORDS    adipocyte protein; cell determination; cell differentiation;
            helix-loop-helix protein.
SOURCE      Rattus sp. cDNA to mRNA.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 3115)
  AUTHORS   Kim,J., Tontonoz,P. and Spiegelman,B.
  TITLE     Add1: a novel helix-loop-helix protein associated with adipocyte
            determination and differentiation
  JOURNAL   Mol. Cell. Biol. 13, 4753-4759 (1993)
  MEDLINE   93330269
FEATURES             Location/Qualifiers
     source          1. .3115
                     /organism="Rattus sp."
                     /db_xref="taxon:10118"
BASE COUNT      629 a   1002 c    870 g    614 t
ORIGIN
```

```
  Query Match             1.5%;  Score 52.6;  DB 12;  Length 3115;
  Best Local Similarity   55.9%;  Pred. No. 0.0011;
  Matches  100;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy   1086 gagcgtcgtacggcacacaatctcatcgaaaagaagtatagatgctcaataaatgatcga 1145
          |||   || || || ||||||  ||| || |||   ||  | | ||| || ||||| |
Db    922 GAGAAGCGCACAGCCCACAATGCCATTGAGAAGCGCTACCGTTCCTCTATCAATGACAAG 981

Qy   1146 attcaacagctgaaagtacttttgtgtggggatgaagctaagctttcaaaatcggcaaca 1205
          |||    |||| || |  ||  ||   ||   ||| || |||||    ||||| ||
Db    982 ATTGTGGAGCTCAAGGACCTGGTGGTGGGCACTGAGGCAAAGCTGAATAAATCTGCTGTC 1041

Qy   1206 ctacgacgggctattgaacatatcgaggaggttgaacacgagaatcaggtgttgaagca 1264
           | ||   ||| || ||  | |||       |  | |||   || |||    | |||||
Db   1042 TTGCGCAAGGCCATCGACTACATCCGCTTCTTACAGCACAGCAACCAGAAACTCAAGCA 1100
```

```
RESULT  15
AR069298
LOCUS       AR069298     4154 bp    DNA             PAT       18-FEB-2000
DEFINITION  Sequence 37 from patent US 5891631.
ACCESSION   AR069298
VERSION     AR069298.1  GI:7220186
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4154)
  AUTHORS   Goldstein,J.L., Brown,M.S., Briggs,M.R. and Wang,X.
  TITLE     Methods relating tosterol regulatory element binding proteins
  JOURNAL   Patent: US 5891631-A 37 06-APR-1999;
FEATURES             Location/Qualifiers
     source          1. .4154
                     /organism="unknown"
BASE COUNT      714 a   1406 c   1280 g    754 t
ORIGIN
```

```
  Query Match             1.5%;  Score 51;  DB 5;  Length 4154;
  Best Local Similarity   51.5%;  Pred. No. 0.0033;
  Matches  117;  Conservative    0;  Mismatches  110;  Indels    0;  Gaps    0;

Qy   1086 gagcgtcgtacggcacacaatctcatcgaaaagaagtatagatgctcaataaatgatcga 1145
          |||   || || || |||||   ||| || |||   ||  | | ||| || |||||   |
Db   1133 GAGAAGCGCACAGCCCACAACGCCATTGAGAAGCGCTACCGCTCCTCCATCAATGACAAA 1192

Qy   1146 attcaacagctgaaagtacttttgtgtggggatgaagctaagctttcaaaatcggcaaca 1205
          ||     |||| || |  ||  ||   ||   ||| || |||||    ||||| ||
```

```
Db    1193 ATCATTGAGCTCAAGGATCTGGTGGTGGGCACTGAGGCAAAGCTGAATAAATCTGCTGTC 1252

Qy    1206 ctacgacgggctattgaacatatcgaggaggttgaacacgagaatcaggtgttgaagcat 1265
           | ||   ||| || ||  | ||       |  |||||   || |||    | |||||
Db    1253 TTGCGCAAGGCCATCGACTACATTCGCTTTCTGCAACACAGCAACCAGAAACTCAAGCAG 1312

Qy    1266 catgttgaacaaatgagaaagacactgcagaataatcgattaccgta 1312
           |      |    || | |   |  | || || |    ||  | | |
Db    1313 GAGAACCTAAGTCTGCGCACTGCTGTCCACAAAAGCAAATCTCTGAA 1359


Search completed: September  3, 2000, 23:42:35
Job time: 18375 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2000, 18:57:10 ; Search time 183.89 Seconds
(without alignments)
2557.526 Million cell updates/sec

Title: US-09-332-522B-1
Perfect score: 3419
Sequence: 1 ggtttaattacccaagtttg.....taaatacgattatatataaa 3419

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	71.8	2.1	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	51	1.5	4154	1	US-08-131-365B-37	Sequence 37, Appl
	3	51	1.5	4154	3	US-08-668-123-37	Sequence 37, Appl
	4	47.6	1.4	5197	1	US-08-131-365B-53	Sequence 53, Appl
	5	47.6	1.4	5197	3	US-08-668-123-53	Sequence 53, Appl
	6	37.4	1.1	1981	5	US-08-747-574-1	Sequence 1, Appli
	7	37.4	1.1	2295	1	US-08-375-300-3	Sequence 3, Appli
	8	37.4	1.1	2295	5	US-09-177-431-3	Sequence 3, Appli
	9	37.4	1.1	2295	6	PCT-US95-16930-3	Sequence 3, Appli
	10	37.4	1.1	4080	1	US-08-375-300-1	Sequence 1, Appli
	11	37.4	1.1	4080	5	US-09-177-431-1	Sequence 1, Appli
	12	37.4	1.1	4080	6	PCT-US95-16930-1	Sequence 1, Appli
	13	36.4	1.1	1027	5	US-08-463-903-19	Sequence 19, Appl
c	14	36	1.1	10747	4	US-08-147-777-1	Sequence 1, Appli
c	15	36	1.1	10747	5	US-08-452-872-1	Sequence 1, Appli
c	16	36	1.1	10747	6	PCT-US93-03985-1	Sequence 1, Appli
c	17	36	1.1	24979	4	US-08-147-777-3	Sequence 3, Appli
c	18	36	1.1	24979	5	US-08-452-872-3	Sequence 3, Appli
c	19	36	1.1	24979	6	PCT-US93-03985-3	Sequence 3, Appli
	20	35.8	1.0	7218	1	US-08-232-463-14	Sequence 14, Appl
	21	34.4	1.0	2344	5	US-08-893-852A-2	Sequence 2, Appli
c	22	34.4	1.0	53526	5	US-08-658-136-2	Sequence 2, Appli
c	23	34.4	1.0	53577	5	US-08-658-136-1	Sequence 1, Appli
	24	33.4	1.0	3489	4	US-08-728-323A-1	Sequence 1, Appli
c	25	33.4	1.0	32207	3	US-08-770-379-20	Sequence 20, Appl
	26	32.8	1.0	2277	2	US-08-676-967-2	Sequence 2, Appli
	27	32.8	1.0	2277	2	US-08-676-974-2	Sequence 2, Appli
	28	32.8	1.0	2277	4	US-09-098-487-2	Sequence 2, Appli
	29	32.8	1.0	2721	7	5215881-2	Patent No. 5215881
	30	32.8	1.0	8438	1	US-07-945-283-1	Sequence 1, Appli
	31	32.6	1.0	570	1	US-07-885-970A-10	Sequence 10, Appl
	32	32.6	1.0	570	1	US-08-298-687A-10	Sequence 10, Appl
	33	32.6	1.0	570	1	US-08-298-829-10	Sequence 10, Appl
c	34	32.6	1.0	609	1	US-08-530-797-9	Sequence 9, Appli
c	35	32.6	1.0	609	4	US-08-787-335-9	Sequence 9, Appli
c	36	32.6	1.0	3680	2	US-08-759-848-1	Sequence 1, Appli
c	37	32.6	1.0	3680	6	PCT-US95-09383-1	Sequence 1, Appli
	38	32.4	0.9	512	2	US-08-460-739-2	Sequence 2, Appli
c	39	32.4	0.9	2188	1	US-07-865-662F-10	Sequence 10, Appl
	40	32	0.9	2085	1	US-08-371-930-9	Sequence 9, Appli
	41	32	0.9	2085	6	PCT-US94-01712-9	Sequence 9, Appli
	42	32	0.9	2088	3	US-08-104-165-4	Sequence 4, Appli
	43	32	0.9	2088	7	5218100-4	Patent No. 5218100
	44	32	0.9	2256	4	US-08-422-333-1	Sequence 1, Appli
	45	32	0.9	2256	7	5187153-1	Patent No. 5187153

ALIGNMENTS

```
RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
;   APPLICANT: DORNER, F.
;   APPLICANT: SCHEIFLINGER, F.
;   APPLICANT: FALKNER, F. G.
;   TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
;   NUMBER OF SEQUENCES: 52
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Foley & Lardner
;     STREET: 1800 Diagonal Road, Suite 500
;     CITY: Alexandria
;     STATE: VA
;     COUNTRY: USA
;     ZIP: 22313-0299
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/232,463
;     FILING DATE:
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/07/935,313
;     FILING DATE:
;     APPLICATION NUMBER: EP 91 114 300.6
;     FILING DATE: 26-AUG-1991
;   ATTORNEY/AGENT INFORMATION:
;     NAME: BENT, Stephen A.
;     REGISTRATION NUMBER: 29,768
;     REFERENCE/DOCKET NUMBER: 30472/114 IMMU
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (703)836-9300
;     TELEFAX: (703)683-4109
;     TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 7218 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;     CLONE: pTZgpt-Fls
US-08-232-463-14
```

```
  Query Match             2.1%;  Score 71.8;  DB 1;  Length 7218;
  Best Local Similarity   11.7%;  Pred. No. 1.2e-11;
  Matches   58;  Conservative  231;  Mismatches  208;  Indels    0;  Gaps    0;

Qy    771 ccatcgattgaagatgctccagagacgaaaccaactcatttggttgaaccacaaagtcca 830
          | ||| ||  |||  | |  |||||   ||  |  |  |  ::  :::  : ::::   :
Db   1475 CTATCTATGCAAGTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRR 1416

Qy    831 aaaagcccgcagaatatgaaagaggagcttcttcggttactagttaacatgtctccgagt 890
          :::::   : :::: : :::::::::       ::  :  ::  :: : :     :::
Db   1415 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1356

Qy    891 gaagttgaacggttaaagaataaaaaatcaggagcatgttcagcgacgaatgggccatcg 950
          ::::  ::: ::  :::::: :::::::  ::::: : :   :: :: ::: :::  :  :
Db   1355 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1296

Qy    951 aggagtaaggagaaggcggcgaagattgtgattcaggagacagcggaaggggatgaagat 1010
          ::::: :::::::::: :: :::::  : ::   :::::: :: :::::::::: :::::
Db   1295 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1236

Qy   1011 gaggatgatgaggatagtgattccggggagactatgtctcagggaactactattattgtt 1070
          ::::: :: ::::: :: ::    :::::::  : :    ::::::  :  :  :  :
Db   1235 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1176

Qy   1071 cgaagaccaaaaaccgagcgtcgtacggcacacaatctcatcgaaaagaagtatagatgc 1130
           :::::  :::::  ::: :  : : :: : : ::    :  :::::::::: : ::: :
Db   1175 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1116

Qy   1131 tcaataaatgatcgaattcaacagctgaaagtacttttgtgtggggatgaagctaagctt 1190
            :: ::: ::  :::   :: ::  ::::: :     : : ::::: :|    |||||
Db   1115 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATCGCAAGCTC 1056

Qy   1191 tcaaaatcggcaacactacgacgggctattgaacatatcgaggaggttgaacacgagaat 1250
           |   | | ||| |    |   |   || |      | ||    |       | || ||
Db   1055 CCTCGACCTGCAGCCAAGCTCGGAATTAATTCTGTGAGCGTATGGCAAACGAAGGAAAAA 996

Qy   1251 caggtgttgaagcatca 1267
          || | | | |||  ||
Db    995 TAGTTATAGTAGCCGCA 979


RESULT   2
US-08-131-365B-37
; Sequence 37, Application US/08131365B
; Patent No. 5527690
;  GENERAL INFORMATION:
;    APPLICANT:  Brown, Michael S.
;    APPLICANT:  Briggs, Michael R.
;    APPLICANT:  Wang, Xiaodong
;    APPLICANT:  Goldstein, Joseph L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS RELATING
;    TITLE OF INVENTION:  TO STEROL REGULATORY ELEMENT BINDING
;    TITLE OF INVENTION:  PROTEINS
;    NUMBER OF SEQUENCES:  64
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  U.S.A.
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/131,365B
;      FILING DATE:  01-OCT-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  UTSD:372/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4154 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "DNA"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  167..3607
US-08-131-365B-37


  Query Match             1.5%;  Score 51;  DB 1;  Length 4154;
  Best Local Similarity   51.5%;  Pred. No. 2.4e-05;
  Matches  117;  Conservative    0;  Mismatches  110;  Indels    0;  Gaps    0;

Qy   1086 gagcgtcgtacggcacacaatctcatcgaaaagaagtatagatgctcaataaatgatcga 1145
          |||   || || || |||||   ||| || |||   ||  | | ||| || |||||   |
Db   1133 GAGAAGCGCACAGCCCACAACGCCATTGAGAAGCGCTACCGCTCCTCCATCAATGACAAA 1192

Qy   1146 attcaacagctgaaagtacttttgtgtggggatgaagctaagctttcaaaatcggcaaca 1205
          ||     |||| || |  ||  ||   ||   ||| || |||||    ||||| ||
Db   1193 ATCATTGAGCTCAAGGATCTGGTGGTGGGCACTGAGGCAAAGCTGAATAAATCTGCTGTC 1252

Qy   1206 ctacgacgggctattgaacatatcgaggaggttgaacacgagaatcaggtgttgaagcat 1265
           | ||   ||| || ||  | ||         |  |||||   || |||    | |||||
Db   1253 TTGCGCAAGGCCATCGACTACATTCGCTTTCTGCAACACAGCAACCAGAAACTCAAGCAG 1312

Qy   1266 catgttgaacaaatgagaaagacactgcagaataatcgattaccgta 1312
           |      |    || | |   |  | || || |    ||  | | |
Db   1313 GAGAACCTAAGTCTGCGCACTGCTGTCCACAAAAGCAAATCTCTGAA 1359


RESULT   3
US-08-668-123-37
; Sequence 37, Application US/08668123
; Patent No. 5891631
;  GENERAL INFORMATION:
;    APPLICANT:  Brown, Michael S.
;    APPLICANT:  Briggs, Michael R.
;    APPLICANT:  Wang, Xiaodong
;    APPLICANT:  Goldstein, Joseph L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS RELATING
;    TITLE OF INVENTION:  TO STEROL REGULATORY ELEMENT BINDING
;    TITLE OF INVENTION:  PROTEINS
;    NUMBER OF SEQUENCES:  64
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  U.S.A.
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/668,123
;      FILING DATE:  14-JUN-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/131,365
;      FILING DATE:  01-OCT-1993
;    ATTORNEY/AGENT INFORMATION:
```

```
;       NAME:  Parker, David L.
;       REGISTRATION NUMBER:  32,165
;       REFERENCE/DOCKET NUMBER:  UTSD:372/PAR
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (512) 418-3000
;       TELEFAX:  (512) 474-7577
;   INFORMATION FOR SEQ ID NO:  37:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4154 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  other nucleic acid
;       DESCRIPTION:  /desc = "DNA"
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  167..3607
US-08-668-123-37


  Query Match             1.5%;  Score 51;  DB 3;  Length 4154;
  Best Local Similarity   51.5%;  Pred. No. 2.4e-05;
  Matches  117;  Conservative    0;  Mismatches  110;  Indels    0;  Gaps    0;

Qy   1086 gagcgtcgtacggcacacaatctcatcgaaaagaagtatagatgctcaataaatgatcga 1145
          |||   || || || |||||   ||| || |||   ||  | | |||| || |||||   |
Db   1133 GAGAAGCGCACAGCCCACAACGCCATTGAGAAGCGCTACCGCTCCTCCATCAATGACAAA 1192

Qy   1146 attcaacagctgaaagtacttttgtgtggggatgaagctaagctttcaaaatcggcaaca 1205
          ||     |||| || |  ||  ||   ||   ||| || ||||    |||| ||
Db   1193 ATCATTGAGCTCAAGGATCTGGTGGTGGGCACTGAGGCAAAGCTGAATAAATCTGCTGTC 1252

Qy   1206 ctacgacgggctattgaacatatcgaggaggttgaacacgagaatcaggtgttgaagcat 1265
            | ||   ||| || ||  | ||        |  |||||   || |||    | |||||
Db   1253 TTGCGCAAGGCCATCGACTACATTCGCTTTCTGCAACACAGCAACCAGAAACTCAAGCAG 1312

Qy   1266 catgttgaacaaatgagaaagacactgcagaataatcgattaccgta 1312
            |      |    || | |   |  | || || |    ||  | | |
Db   1313 GAGAACCTAAGTCTGCGCACTGCTGTCCACAAAAGCAAATCTCTGAA 1359


RESULT   4
US-08-131-365B-53
; Sequence 53, Application US/08131365B
; Patent No. 5527690
;  GENERAL INFORMATION:
;    APPLICANT:  Brown, Michael S.
;    APPLICANT:  Briggs, Michael R.
;    APPLICANT:  Wang, Xiaodong
;    APPLICANT:  Goldstein, Joseph L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS RELATING
;    TITLE OF INVENTION:  TO STEROL REGULATORY ELEMENT BINDING
;    TITLE OF INVENTION:  PROTEINS
;    NUMBER OF SEQUENCES:  64
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  U.S.A.
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/131,365B
;      FILING DATE:  01-OCT-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  UTSD:372/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  53:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5197 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "DNA"
US-08-131-365B-53


  Query Match             1.4%;  Score 47.6;  DB 1;  Length 5197;
  Best Local Similarity   54.6%;  Pred. No. 0.00032;
  Matches   95;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy   1079 aaaaaccgagcgtcgtacggcacacaatctcatcgaaaagaagtatagatgctcaataaa 1138
          | ||   ||  | || ||  | || ||| |||| || ||    ||| | | ||| || ||
Db   1098 AGAAGGAGAAAGGCGGACAACCCATAATATCATTGAGAAACGATATCGCTCCTCCATCAA 1157

Qy   1139 tgatcgaattcaacagctgaaagtacttttgtgtggggatgaagctaagctttcaaaatc 1198
          |||   |||      |  |||||  || |    |||   || || ||| |    || ||
Db   1158 TGACAAAATCATCGAATTGAAAGACCTGGTCATGGGGACAGACGCCAAGATGCACAAGTC 1217

Qy   1199 ggcaacactacgacgggctattgaacatatcgaggaggttgaacacgagaatca 1252
            |     ||  |   ||| |||||  | ||| |  |  |  | || |  ||||
Db   1218 TGGCGTTCTGAGGAAGGCCATTGATTACATCAAATACTTGCAGCAGGTCAATCA 1271


RESULT   5
US-08-668-123-53
; Sequence 53, Application US/08668123
; Patent No. 5891631
;  GENERAL INFORMATION:
;    APPLICANT:  Brown, Michael S.
;    APPLICANT:  Briggs, Michael R.
;    APPLICANT:  Wang, Xiaodong
;    APPLICANT:  Goldstein, Joseph L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS RELATING
;    TITLE OF INVENTION:  TO STEROL REGULATORY ELEMENT BINDING
;    TITLE OF INVENTION:  PROTEINS
;    NUMBER OF SEQUENCES:  64
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  U.S.A.
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/668,123
;      FILING DATE:  14-JUN-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/131,365
;      FILING DATE:  01-OCT-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  UTSD:372/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  53:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  5197 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "DNA"
US-08-668-123-53


  Query Match             1.4%;  Score 47.6;  DB 3;  Length 5197;
  Best Local Similarity   54.6%;  Pred. No. 0.00032;
  Matches   95;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy    1079 aaaaaccgagcgtcgtacggcacacaatctcatcgaaaagaagtatagatgctcaataaa 1138
           | ||   ||  | || ||  | || ||| |||| || ||    ||| | | ||| || ||
Db    1098 AGAAGGAGAAAGGCGGACAACCCATAATATCATTGAGAAACGATATCGCTCCTCCATCAA 1157

Qy    1139 tgatcgaattcaacagctgaaagtacttttgtgtggggatgaagctaagctttcaaaatc 1198
           |||   |||      |  |||||  ||  |    |||   || || ||| |    || ||
Db    1158 TGACAAAATCATCGAATTGAAAGACCTGGTCATGGGGACAGACGCCAAGATGCACAAGTC 1217

Qy    1199 ggcaacactacgacgggctattgaacatatcgaggaggttgaacacgagaatca 1252
           |     ||  |   ||| |||||  | ||| |  |  |  | || |  |||||
Db    1218 TGGCGTTCTGAGGAAGGCCATTGATTACATCAAATACTTGCAGCAGGTCAATCA 1271


RESULT   6
US-08-747-574-1
; Sequence 1, Application US/08747574
; Patent No. 6015939
;  GENERAL INFORMATION:
;    APPLICANT:  CALGENE, INC.
;    TITLE OF INVENTION:  PLANT VDE GENES AND
;    TITLE OF INVENTION:  METHODS RELATED THERETO
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Calgene, Inc.
;      STREET:  1920 Fifth Street
;      CITY:  Davis
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  95616
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.40 MB
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh 7.5
;      SOFTWARE:  Microsoft Word 5.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/747,574
;      FILING DATE:  No. 6015939ember 7, 1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/023,502
;      FILING DATE:  August 6, 1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/006,315
;      FILING DATE:  No. 6015939ember 7, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Donna E. Scherer
;      REGISTRATION NUMBER:  34,719
;      NAME:  Carl J. Schwedler
;      REGISTRATION NUMBER:  36,924
;      REFERENCE/DOCKET NUMBER:  119-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (916) 753-6313
;      TELEFAX:  (916) 753-1510
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1981
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
US-08-747-574-1


  Query Match             1.1%;  Score 37.4;  DB 5;  Length 1981;
  Best Local Similarity   62.1%;  Pred. No. 0.24;
  Matches   59;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy     960 gagaaggcggcgaagattgtgattcaggagacagcggaaggggatgaagatgaggatgat 1019
           ||||||  |  ||  |  |    |  |||| ||  |||| || |||| | ||||  ||
Db    1405 GAGAAGTTGTTGATAAAAGAAGCTGTAGAGATAGAAGAAGAGGTTGAAAAAGAGGTGGAG 1464

Qy    1020 gaggatagtgattccggggagactatgtctcaggg 1054
           ||| ||| ||| | | |  |||| ||| |||| |
Db    1465 AAGGTTAGAGATACTGAGATGACTTTGTTTCAGAG 1499


RESULT   7
US-08-375-300-3
; Sequence 3, Application US/08375300
; Patent No. 5679566
;  GENERAL INFORMATION:
;    APPLICANT:  Feng, He
;    APPLICANT:  Jacobson, Allan S.
;    TITLE OF INVENTION:  HETEROLOGOUS POLYPEPTIDE PRODUCTION IN
;    TITLE OF INVENTION:  THE ABSENCE OF NONSENSE-MEDIATED mRNA DECAY FUNCTION
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street Suite 3100
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/375,300
;      FILING DATE:  20-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. P.
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  04020/046001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)542-5070
;      TELEFAX:  (617)542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2295 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-375-300-3


  Query Match             1.1%;  Score 37.4;  DB 1;  Length 2295;
  Best Local Similarity   53.0%;  Pred. No. 0.27;
  Matches   80;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy     889 gtgaagttgaacggttaaagaataaaaaatcaggagcatgttcagcgacgaatgggccat 948
           |||  | |  ||    | |  || || |  | |    | || || |||  |  | | ||
Db    1589 GTGTCGATTTACTAGGAGAAGATGAAGACGCGGAGATAAGTACACCGAACACAGAGTCAG 1648

Qy     949 cgaggagtaaggagaaggcggcgaagattgtgattcaggagacagcggaaggggatgaag 1008
           ||    | ||  |  ||||   | |    |  | | | ||   ||  || | ||| || |
Db    1649 CGCCAGGAAAACATCAGGCAAAGCAAGACGAAAGTGAAGATGAAGACGATGAGGACGATG 1708

Qy    1009 atgaggatgatgaggatagtgattccgggga 1039
           | |||||||||||| |||   |||  ||  ||
```

```
Db   1709 ACGAGGATGATGACGATGACGATGACGATGA 1739


RESULT   8
US-09-177-431-3
; Sequence 3, Application US/09177431
; Patent No. 6071700
;  GENERAL INFORMATION:
;    APPLICANT:  He, Feng
;    APPLICANT:  Jacobson, Allan S.
;    TITLE OF INVENTION:  HETEROLOGOUS POLYPEPTIDE PRODUCTION IN THE
;    TITLE OF INVENTION:  ABSENCE OF NONSENSE-MEDIATED mRNA DECAY FUNCTION
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/177,431
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/955,472
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. Peter
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  07917/050001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-9806
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2295 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-177-431-3


  Query Match             1.1%;  Score 37.4;  DB 5;  Length 2295;
  Best Local Similarity   53.0%;  Pred. No. 0.27;
  Matches   80;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy    889 gtgaagttgaacggttaaagaataaaaaatcaggagcatgttcagcgacgaatgggccat 948
          |||  | |  ||    | |  || || |  | |    | || || |||  |  | | ||
Db   1589 GTGTCGATTTACTAGGAGAAGATGAAGACGCGGAGATAAGTACACCGAACACAGAGTCAG 1648

Qy    949 cgaggagtaaggagaaggcggcgaagattgtgattcaggagacagcggaaggggatgaag 1008
          ||    | ||  |  ||||   | |    |  | | | ||   ||  || | ||| || |
Db   1649 CGCCAGGAAAACATCAGGCAAAGCAAGACGAAAGTGAAGATGAAGACGATGAGGACGATG 1708

Qy   1009 atgaggatgatgaggatagtgattccgggga 1039
          | ||||||||||| |||   |||  ||  ||
Db   1709 ACGAGGATGATGACGATGACGATGACGATGA 1739


RESULT   9
PCT-US95-16930-3
; Sequence 3, Application PC/TUS9516930
;  GENERAL INFORMATION:
;    APPLICANT:  UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL
;    TITLE OF INVENTION:  HETEROLOGOUS POLYPEPTIDE
;    TITLE OF INVENTION:  PRODUCTION IN THE ABSENCE OF
;    TITLE OF INVENTION:  NONSENSE-MEDIATED mRNA DECAY
;    TITLE OF INVENTION:  FUNCTION
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street Suite 3100
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/16930
;      FILING DATE:  27-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/375,300
;      FILING DATE:  20-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. Peter
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  04020/046WO1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)542-5070
;      TELEFAX:  (617)542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2295 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
PCT-US95-16930-3


  Query Match             1.1%;  Score 37.4;  DB 6;  Length 2295;
  Best Local Similarity   53.0%;  Pred. No. 0.27;
  Matches   80;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy    889 gtgaagttgaacggttaaagaataaaaaatcaggagcatgttcagcgacgaatgggccat 948
          |||  | |  ||    | |  || || |  | |    | || || |||  |  | | ||
Db   1589 GTGTCGATTTACTAGGAGAAGATGAAGACGCGGAGATAAGTACACCGAACACAGAGTCAG 1648

Qy    949 cgaggagtaaggagaaggcggcgaagattgtgattcaggagacagcggaaggggatgaag 1008
          ||    | ||  |  ||||   | |    |  | | | ||   ||  || | ||| || |
Db   1649 CGCCAGGAAAACATCAGGCAAAGCAAGACGAAAGTGAAGATGAAGACGATGAGGACGATG 1708

Qy   1009 atgaggatgatgaggatagtgattccgggga 1039
          | ||||||||||| |||   |||  ||  ||
Db   1709 ACGAGGATGATGACGATGACGATGACGATGA 1739


RESULT  10
US-08-375-300-1
; Sequence 1, Application US/08375300
; Patent No. 5679566
;  GENERAL INFORMATION:
;    APPLICANT:  Feng, He
;    APPLICANT:  Jacobson, Allan S.
;    TITLE OF INVENTION:  HETEROLOGOUS POLYPEPTIDE PRODUCTION IN
;    TITLE OF INVENTION:  THE ABSENCE OF NONSENSE-MEDIATED mRNA DECAY FUNCTION
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street Suite 3100
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/375,300
;      FILING DATE:  20-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. P.
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  04020/046001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)542-5070
;      TELEFAX:  (617)542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4080 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-375-300-1


  Query Match             1.1%;  Score 37.4;  DB 1;  Length 4080;
  Best Local Similarity   53.0%;  Pred. No. 0.39;
  Matches   80;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy    889 gtgaagttgaacggttaaagaataaaaaatcaggagcatgttcagcgacgaatgggccat 948
          |||  | |  ||    | |  || || |  | |    | || || |||  |  | | ||
Db   3038 GTGTCGATTTACTAGGAGAAGATGAAGACGCGGAGATAAGTACACCGAACACAGAGTCAG 3097

Qy    949 cgaggagtaaggagaaggcggcgaagattgtgattcaggagacagcggaaggggatgaag 1008
          ||    | ||  |  ||||   | |    |  | | | ||   ||  || | ||| || |
Db   3098 CGCCAGGAAAACATCAGGCAAAGCAAGACGAAAGTGAAGATGAAGACGATGAGGACGATG 3157

Qy   1009 atgaggatgatgaggatagtgattccgggga 1039
          | ||||||||||| |||   |||  ||  ||
Db   3158 ACGAGGATGATGACGATGACGATGACGATGA 3188


RESULT  11
US-09-177-431-1
; Sequence 1, Application US/09177431
; Patent No. 6071700
;  GENERAL INFORMATION:
;    APPLICANT:  He, Feng
;    APPLICANT:  Jacobson, Allan S.
;    TITLE OF INVENTION:  HETEROLOGOUS POLYPEPTIDE PRODUCTION IN THE
;    TITLE OF INVENTION:  ABSENCE OF NONSENSE-MEDIATED mRNA DECAY FUNCTION
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/177,431
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/955,472
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. Peter
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  07917/050001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-9806
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4080 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-177-431-1


  Query Match             1.1%;  Score 37.4;  DB 5;  Length 4080;
  Best Local Similarity   53.0%;  Pred. No. 0.39;
  Matches   80;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy    889 gtgaagttgaacggttaaagaataaaaaatcaggagcatgttcagcgacgaatgggccat 948
          |||  | |  ||    | |  || || |  | |    | || || |||  |  | | ||
Db   3038 GTGTCGATTTACTAGGAGAAGATGAAGACGCGGAGATAAGTACACCGAACACAGAGTCAG 3097

Qy    949 cgaggagtaaggagaaggcggcgaagattgtgattcaggagacagcggaaggggatgaag 1008
          ||    | ||  |  ||||   | |    |  | | | ||   ||  || | ||| || |
Db   3098 CGCCAGGAAAACATCAGGCAAAGCAAGACGAAAGTGAAGATGAAGACGATGAGGACGATG 3157

Qy   1009 atgaggatgatgaggatagtgattccgggga 1039
          | ||||||||||| |||   |||  ||  ||
Db   3158 ACGAGGATGATGACGATGACGATGACGATGA 3188


RESULT  12
PCT-US95-16930-1
; Sequence 1, Application PC/TUS9516930
;  GENERAL INFORMATION:
;    APPLICANT:  UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL
;    TITLE OF INVENTION:  HETEROLOGOUS POLYPEPTIDE
;    TITLE OF INVENTION:  PRODUCTION IN THE ABSENCE OF
;    TITLE OF INVENTION:  NONSENSE-MEDIATED mRNA DECAY
;    TITLE OF INVENTION:  FUNCTION
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street Suite 3100
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/16930
;      FILING DATE:  27-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/375,300
;      FILING DATE:  20-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. Peter
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  04020/046WO1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)542-5070
;      TELEFAX:  (617)542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4080 base pairs
;      TYPE:  nucleic acid
```

```
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
PCT-US95-16930-1


  Query Match             1.1%;  Score 37.4;  DB 6;  Length 4080;
  Best Local Similarity   53.0%;  Pred. No. 0.39;
  Matches   80;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy    889 gtgaagttgaacggttaaagaataaaaaatcaggagcatgttcagcgacgaatgggccat 948
          ||| | | ||    | |  || || | | |    | || || ||| |  | | ||
Db   3038 GTGTCGATTTACTAGGAGAAGATGAAGACGCGGAGATAAGTACACCGAACACAGAGTCAG 3097

Qy    949 cgaggagtaaggagaaggcggcgaagattgtgattcaggagacagcggaaggggatgaag 1008
          ||    | ||  |  ||||   | |    |  | | | ||   ||  || | ||| || |
Db   3098 CGCCAGGAAAACATCAGGCAAAGCAAGACGAAAGTGAAGATGAAGACGATGAGGACGATG 3157

Qy   1009 atgaggatgatgaggatagtgattccgggga 1039
          | ||||||||||| |||   |||  ||  ||
Db   3158 ACGAGGATGATGACGATGACGATGACGATGA 3188


RESULT  13
US-08-463-903-19
; Sequence 19, Application US/08463903
; Patent No. 6071515
; GENERAL INFORMATION:
;  APPLICANT: Mezes, Peter S.
;  APPLICANT: Richard, Ruth A.
;  APPLICANT: Affholter, Joseph A.
;  APPLICANT: Kotite, Nicolas J.
;  TITLE OF INVENTION: Dimer and Multimer Forms of Single Chain Polypeptides
;  FILE REFERENCE: 40224A US
;  CURRENT APPLICATION NUMBER: US/08/463,903
;  CURRENT FILING DATE: 1995-06-05
;  EARLIER APPLICATION NUMBER: US 07/935,695
;  EARLIER FILING DATE: 1992-08-21
;  NUMBER OF SEQ ID NOS: 102
;  SOFTWARE: MS-Word for Windows, Ver. 7.0
; SEQ ID NO 19
;   LENGTH: 1027
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: Hum4 VL-UNIHOPE linker-FLAG peptide construct in pATDFLAG
;   LOCATION: 1..1027
US-08-463-903-19


  Query Match             1.1%;  Score 36.4;  DB 5;  Length 1027;
  Best Local Similarity   54.7%;  Pred. No. 0.32;
  Matches   93;  Conservative    0;  Mismatches   76;  Indels    1;  Gaps    1;

Qy    969 gcgaagattgtgattcaggagacagcggaaggggatgaagatgaggatgatgaggatagt 1028
          ||| |   || ||   ||||  | ||| |   |||||| | | || | || || | ||
Db    704 gcggacgatgcgaaaaaggatgctgcgaagaaggatgacgctaagaaagacgatgctaaa 763

Qy   1029 gattccggggagactatgtctcagggaactactattattgt-tcgaagaccaaaaaccga 1087
          |   |  ||||| ||||| |  |  ||||| |  |   |   ||  |  ||||||| |
Db    764 aaggacctcgagacaatgtcgctagcgactacaaggacgatgatgacaaataaaaaccta 823

Qy   1088 gcgtcgtacggcacacaatctcatcgaaaagaagtatagatgctcaataa 1137
          |||  | |     || ||  |||||  | | ||    |  || | |  ||
Db    824 gcgatgaatccgtcaaaacatcatcttacataaagtcacttggtgatcaa 873


RESULT  14
US-08-147-777-1/c
; Sequence 1, Application US/08147777
; Patent No. 5914265
;  GENERAL INFORMATION:
;    APPLICANT:  Roop, Dennis R.
;    APPLICANT:  Rothnagel, Joseph A.
;    APPLICANT:  Greenhalgh, David A.
;    APPLICANT:  Yuspa, Stuart H.
;    TITLE OF INVENTION:  KERATIN K1 EXPRESSION VECTORS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LYON & LYON
;      STREET:  611 West Sixth Street
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90017
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  IBM MS-DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/147,777
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;    PRIOR APPLICATION DATA:  including application
;    PRIOR APPLICATION DATA:  described below:             two
;      APPLICATION NUMBER:  07/876,289
;      FILING DATE:  April 30, 1992
;      APPLICATION NUMBER:  Unassigned (204/144)
;      FILING DATE:  October 29, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  204/153
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10747 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-147-777-1


  Query Match             1.1%;  Score 36;  DB 4;  Length 10747;
  Best Local Similarity   48.5%;  Pred. No. 2.1;
  Matches   99;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy    637 tgtctccatacgaccaagctcaaggcccatcaggaccatcatattactcacaacaccatc 696
          || || | |     | | |   ||| |   |||||||||| |     | || | |  | |
Db   1771 TGACTTCTTGTATGCCACCAGGAGGGCAGACAGGACCATAACCACCCCCATATCCACCAC 1712

Qy    697 aatctccaccacctcatcaccaccatcaccacccgatgccaaaaatccatgagaaccctg 756
            |    |||||||  |   | |  |  |  ||  ||||||||   |   ||   ||
Db   1711 CCCCAAAGCCACCTCCACCAAAACCACCACCACCACTGCCAAAACCACCAAAGCCACCAC 1652

Qy    757 aacaagtggcatctccatcgattgaagatgctccagagacgaaaccaactcatttggttg 816
            | | || || |||||  | |    |    | ||| || |   ||||  |  |
Db   1651 CCCCAATGCCACCTCCACCAAAGCCACCACCACCAAAGCCACCACCACCAAAGCCACCAC 1592

Qy    817 aaccacaaagtccaaaaagcccgc 840
            |||| ||   |||  ||  || |
Db   1591 CACCATAACCACCACCAAAGCCAC 1568


RESULT  15
US-08-452-872-1/c
; Sequence 1, Application US/08452872
; Patent No. 6057298
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Roop, Dennis R.
;    APPLICANT:  Rothnagel, Joseph A.
;    APPLICANT:  Greenhalgh, David A.
;    APPLICANT:  Yuspa, Stuart H.
;    TITLE OF INVENTION:  KERATIN K1 EXPRESSION VECTORS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LYON & LYON
;      STREET:  611 West Sixth Street
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90017
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  IBM MS-DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/452,872
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/147,777
;      FILING DATE:
;      APPLICATION NUMBER:  07/876,289
;      FILING DATE:  April 30, 1992
;      APPLICATION NUMBER:  Unassigned (204/144)
;      FILING DATE:  October 29, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  204/153
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10747 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-452-872-1


  Query Match             1.1%;  Score 36;  DB 5;  Length 10747;
  Best Local Similarity   48.5%;  Pred. No. 2.1;
  Matches   99;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy     637 tgtctccatacgaccaagctcaaggcccatcaggaccatcatattactcacaacaccatc 696
           || || | |     | | |   ||| |   ||||||||| |     | || | |  |  |
Db    1771 TGACTTCTTGTATGCCACCAGGAGGGCAGACAGGACCATAACCACCCCCATATCCACCAC 1712

Qy     697 aatctccaccacctcatcaccaccatcaccacccgatgccaaaaatccatgagaaccctg 756
             |    |||||||  |   | |  |  |  ||  ||||||||   |   ||   ||
Db    1711 CCCCAAAGCCACCTCCACCAAAACCACCACCACCACTGCCAAAACCACCAAAGCCACCAC 1652

Qy     757 aacaagtggcatctccatcgattgaagatgctccagagacgaaaccaactcatttggttg 816
             | | || || ||||| | |     |     | ||| || |   |||| |  |
Db    1651 CCCCAATGCCACCTCCACCAAAGCCACCACCACCAAAGCCACCACCACCAAAGCCACCAC 1592

Qy     817 aaccacaaagtccaaaaagcccgc 840
           |||| ||   |||  ||  || |
Db    1591 CACCATAACCACCACCAAAGCCAC 1568


Search completed: September  4, 2000, 00:29:28
Job time: 19938 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2000, 19:57:05 ; Search time 299.13 Seconds
(without alignments)
2859.649 Million cell updates/sec

Title: US-09-332-522B-1
Perfect score: 3419
Sequence: 1 ggtttaattacccaagtttg.....taaatacgattatatataaa 3419

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	51	1.5	4154	1	Q79037	Human SREBP-1a cod
	2	47.6	1.4	5197	1	Q79038	Human SREBP-2 codi
	3	43.8	1.3	3060	1	Q03098	Enod2a genomic clo
	4	43.8	1.3	3856	1	Q03097	Enod2b genomic clo
c	5	38.2	1.1	672	1	Q70140	Cadrac cDNA. Agent
	6	37.4	1.1	1981	1	T66241	Romaine lettuce vi
	7	37.4	1.1	2295	1	T31994	Nonsense-mediated
	8	37.4	1.1	4080	1	T31993	Nonsense-mediated
	9	36.6	1.1	289	1	X07421	Borrelia burgdorfe
	10	36.4	1.1	1028	1	Q45606	Sequence of pATDFL
c	11	36.2	1.1	11871	1	X13108	Enterococcus faeca
	12	36	1.1	1648	1	Q25532	Sequence of genomi
c	13	35.6	1.0	114955	1	X53491	Human adenosine A1
c	14	34.8	1.0	2758	1	T11575	Human glucagon deg
	15	34.4	1.0	2344	1	X19798	Human growth regul
c	16	34.4	1.0	53526	1	T94101	Human PKD1 gene. H
c	17	34.4	1.0	53577	1	T18551	Human polycystic k
c	18	34.4	1.0	53577	1	T94108	Human PKD1 locus b
	19	34	1.0	1193	1	X25017	Alfalfa cyclin-dep
	20	33.8	1.0	924	1	V71744	Human V3 loop HIV
	21	33.8	1.0	1596	1	V43619	Human secreted pro
	22	33.8	1.0	1679	1	X00719	Human secreted pro
	23	33.8	1.0	1863	1	X00689	Human secreted pro
c	24	33.8	1.0	4418	1	V74765	Staphylococcus aur
	25	33.4	1.0	215	1	X12235	Human biallelic po
c	26	33.4	1.0	32207	1	V73805	KSHV LUR DNA (nucl
c	27	33.4	1.0	117213	1	V62176	HSV-2 strain SB5 C
c	28	33.4	1.0	137507	1	V19941	KSHV long unique c
	29	33.2	1.0	2417	1	V22222	SIRE-1 cDNA from 1
c	30	33.2	1.0	14141	1	X13331	Enterococcus faeca
c	31	33	1.0	6022	1	X23983	Human hGT1 DNA. A
	32	33	1.0	13856	1	V74342	Staphylococcus aur
	33	32.8	1.0	2277	1	V05370	Human telomerase p
	34	32.8	1.0	2277	1	V13834	Homo sapiens ambig
	35	32.8	1.0	2721	1	Q10212	BamH1 J-I fragment
	36	32.8	1.0	2721	1	Q10543	BamHI J-I fragment
	37	32.8	1.0	8438	1	Q73500	DNA encoding Pseud
	38	32.6	1.0	570	1	T13041	Cotton fibre-speci
	39	32.6	1.0	570	1	T30258	Cotton fibre cell-
	40	32.6	1.0	570	1	T70048	Cotton fibre speci
c	41	32.6	1.0	609	1	T62617	Cotton fibre speci
	42	32.6	1.0	1476	1	T08578	Zinc finger protei
c	43	32.6	1.0	3680	1	T10665	Mouse bradykinin B
	44	32.6	1.0	11739	1	X13087	Enterococcus faeca
	45	32.4	0.9	400	1	V78220	Staphylococcus aur

ALIGNMENTS

RESULT 1
Q79037
ID Q79037 standard; DNA; 4154 BP.
AC Q79037;
DT 04-AUG-1995 (first entry)
DE Human SREBP-1a coding sequence.
KW Probe; sterol regulatory element binding protein; regulatory protein;
KW cholesterol metabolism; sterol regulatory element-1; SREBP; SRE-1;
KW SREBP-1; SREBP-2; basic-helix-loop-helix-leucine zipper; plasma;
KW transcription factor; low density lipoporotein; LDL; receptor;
KW 3-hydroxy-3-methylglutaryl coenzyme A synthase; cholesterol;
KW hypercholesterolaemia; enhancer; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 167. .3610
FT /*tag= a
FT /product= SREBP-1a
PN WO9426922-A.
PD 24-NOV-1994.
PF 13-MAY-1994; U05300.
PR 13-MAY-1993; US-061697.
PR 01-OCT-1993; US-131365.
PA (TEXA) UNIV TEXAS SYSTEM.
PI Briggs MR, Brown MS, Goldstein JL, Wang X;
DR WPI; 95-006813/01.
DR P-PSDB; R66390.
PT New sterol regulator element binding protein - used to develop
PT prods. and screening assays for agents for reducing plasma
PT cholesterol levels (Eng)
PS Claim 31; Page 185-98; 305pp; English.
CC The sequences given in Q79037-38 encode the sterol regulatory element
CC binding proteins (SREBP), SREBP-1a and SREBP-2. SREBP's are regulatory
CC proteins which are involved in the regulation of genes involved in
CC cholesterol metabolism that are under the control of an associated
CC sterol regulatory element-1 (SRE-1) enhancer sequence. SREBP proteins
CC fall into two families, SREBP-1 and SREBP-2. Both proteins are members
CC of a family of basic-helix-loop-helix-leucine zipper (bHLH-Zip)
CC transcription factors. Each have the ability to bind to SRE sequences
CC and modulate SRE-mediated transcription. SRE-1 is a conditional enhancer
CC found in the promoters for the low density lipoporotein (LDL) receptor
CC and 3-hydroxy-3-methylglutaryl coenzyme A synthase genes. It increases
CC transcription in the absence of sterols and is inactivated when sterols
CC accumulate. Human SREBP-2 contins 1141 amino acids and has 47% identity
CC with STREBP-1a. SREBP-1a was the first recognised member of this family
CC and has 1147 amino acids. The resemblence between SREBP-1a and SREBP-2
CC includes an acidic N-terminus, a highly conserved bHLH-Zip motif (71%
CC identical), and an unusually long extension of 740 amino acids on the
CC C-terminal side of the bHLH-Zip region. SREBP-2 posseses one feature
CC lacking in SREBP-1a, a Glu rich region (27% Glu over 121 residues).
CC SREBP promotes SRE-1-mediated gene transcription, eg. LDL receptor
CC production in the presence of sterols. SREBP identified in screening
CC assays, may be used to reduce plasma cholesterol levels and in
CC controlling hypercholesterolaemia and its associated diseases.
SQ Sequence 4154 BP; 714 A; 1406 C; 1280 G; 754 T;

```
  Query Match             1.5%;  Score 51;  DB 1;  Length 4154;
  Best Local Similarity   51.5%;  Pred. No. 3.5e-05;
  Matches  117;  Conservative    0;  Mismatches  110;  Indels    0;  Gaps    0;

Qy   1086 gagcgtcgtacggcacacaatctcatcgaaaagaagtatagatgctcaataaatgatcga 1145
          |||   || || || |||||   ||| || |||   ||  | | ||| || |||||   |
Db   1133 GAGAAGCGCACAGCCCACAACGCCATTGAGAAGCGCTACCGCTCCTCCATCAATGACAAA 1192

Qy   1146 attcaacagctgaaagtacttttgtgtggggatgaagctaagctttcaaaatcggcaaca 1205
          ||     |||| || |  ||  ||   ||   ||| || |||||    ||||| ||
Db   1193 ATCATTGAGCTCAAGGATCTGGTGGTGGGCACTGAGGCAAAGCTGAATAAATCTGCTGTC 1252

Qy   1206 ctacgacgggctattgaacatatcgaggaggttgaacacgagaatcaggtgttgaagcat 1265
          | ||   ||| || ||  | ||         | |||||   || |||    | |||||
Db   1253 TTGCGCAAGGCCATCGACTACATTCGCTTTCTGCAACACAGCAACCAGAAACTCAAGCAG 1312

Qy   1266 catgttgaacaaatgagaaagacactgcagaataatcgattaccgta 1312
            |    |    || | |   |  | || || |    ||  | | |
Db   1313 GAGAACCTAAGTCTGCGCACTGCTGTCCACAAAAGCAAATCTCTGAA 1359


RESULT   2
Q79038
ID   Q79038 standard; DNA; 5197 BP.
AC   Q79038;
DT   04-AUG-1995  (first entry)
DE   Human SREBP-2 coding sequence.
KW   Probe; sterol regulatory element binding protein; regulatory protein;
KW   cholesterol metabolism; sterol regulatory element-1; SREBP; SRE-1;
KW   SREBP-1; SREBP-2; basic-helix-loop-helix-leucine zipper; plasma;
KW   transcription factor; low density lipoporotein; LDL; receptor;
KW   3-hydroxy-3-methylglutaryl coenzyme A synthase; cholesterol;
KW   hypercholesterolaemia; enhancer; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             118. .3543
FT                   /*tag= a
FT                   /product= SREBP-2
PN   WO9426922-A.
PD   24-NOV-1994.
PF   13-MAY-1994; U05300.
PR   13-MAY-1993; US-061697.
PR   01-OCT-1993; US-131365.
PA   (TEXA ) UNIV TEXAS SYSTEM.
PI   Briggs MR,  Brown MS,  Goldstein JL,  Wang X;
DR   WPI; 95-006813/01.
DR   P-PSDB; R66391.
PT   New sterol regulator element binding protein - used to develop
PT   prods. and screening assays for agents for reducing plasma
PT   cholesterol levels (Eng)
PS   Claim 34; Page 219-26; 305pp; English.
CC   The sequences given in Q79037-38 encode the sterol regulatory element
CC   binding proteins (SREBP), SREBP-1a and SREBP-2. SREBP's are regulatory
CC   proteins which are involved in the regulation of genes involved in
CC   cholesterol metabolism that are under the control of an associated
CC   sterol regulatory element-1 (SRE-1) enhancer sequence. SREBP proteins
CC   fall into two families, SREBP-1 and SREBP-2. Both proteins are members
CC   of a family of basic-helix-loop-helix-leucine zipper (bHLH-Zip)
CC   transcription factors. Each have the ability to bind to SRE sequences
CC   and modulate SRE-mediated transcription.  SRE-1 is a conditional enhancer
CC   found in the promoters for the low density lipoporotein (LDL) receptor
CC   and 3-hydroxy-3-methylglutaryl coenzyme A synthase genes.  It increases
CC   transcription in the absence of sterols and is inactivated when sterols
CC   accumulate.  Human SREBP-2 contins 1141 amino acids and has 47% identity
CC   with STREBP-1a.  SREBP-1a was the first recognised member of this family
CC   and has 1147 amino acids.  The resemblence between SREBP-1a and SREBP-2
CC   includes an acidic N-terminus, a highly conserved bHLH-Zip motif (71%
CC   identical), and an unusually long extension of 740 amino acids on the
CC   C-terminal side of the bHLH-Zip region.  SREBP-2 posseses one feature
CC   lacking in SREBP-1a, a Glu rich region (27% Glu over 121 residues).
CC   SREBP promotes SRE-1-mediated gene transcription, eg. LDL receptor
CC   production in the presence of sterols.  SREBP identified in screening
CC   assays, may be used to reduce plasma cholesterol levels and in
CC   controlling hypercholesterolaemia and its associated diseases.
SQ   Sequence   5197 BP;    1066 A;    1570 C;    1434 G;    1127 T;


  Query Match             1.4%;  Score 47.6;  DB 1;  Length 5197;
  Best Local Similarity   54.6%;  Pred. No. 0.00047;
  Matches   95;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy   1079 aaaaaccgagcgtcgtacggcacacaatctcatcgaaaagaagtatagatgctcaataaa 1138
          | ||   ||  | || ||  | || ||| |||| || ||    ||| | | ||| || ||
Db   1098 AGAAGGAGAAAGGCGGACAACCCATAATATCATTGAGAAACGATATCGCTCCTCCATCAA 1157

Qy   1139 tgatcgaattcaacagctgaaagtacttttgtgtggggatgaagctaagctttcaaaatc 1198
          |||   |||     |  |||||| ||  |    |||   || || ||| |    || ||
Db   1158 TGACAAAATCATCGAATTGAAAGACCTGGTCATGGGGACAGACGCCAAGATGCACAAGTC 1217

Qy   1199 ggcaacactacgacgggctattgaacatatcgaggaggttgaacacgagaatca 1252
            |     ||  |  ||| ||||| | ||| |  |  |  | || |  |||||
Db   1218 TGGCGTTCTGAGGAAGGCCATTGATTACATCAAATACTTGCAGCAGGTCAATCA 1271


RESULT   3
Q03098
ID   Q03098 standard; cDNA; 3060 BP.
AC   Q03098;
DT   05-JUN-1990  (first entry)
DE   Enod2a genomic clone.
KW   Enod2; soybean; root nodule; regulatory region; nodulin 75; ss.
OS   Glycine max (L) Merr. cv. Williams.
FH   Key             Location/Qualifiers
FT   misc_rna        520. .1565
FT                   /function=regulatory region
FT                   /*tag= a
FT   cds             1653. .2582
FT                   /*tag= b
FT                   /label=ORF1
FT                   /note="actual coding sequence of Enod2 encoding N-75"
FT   cds             1750. .2469
FT                   /*tag= c
FT                   /label=ORF2
FT   tata_signal     1490. .1494
FT                   /*tag= d
FT   caat_signal     1478. .
FT                   /*tag= e
FT   enhancer        567. .575
FT                   /*tag= f
FT   enhancer        979. .986
FT                   /*tag= g
FT   enhancer        1027. .1034
FT                   /*tag= h
FT   enhancer        1377. .1384
FT                   /*tag= i
FT   enhancer        1404. .1411
FT                   /*tag= j
PN   EP-349338-A.
PD   03-JAN-1990.
PF   30-JUN-1989; 306666.
PR   01-JUL-1988; US-214297.
PA   (LUBR) Lubrizol Genetics.
PI   Franssen H, Bisseling AH;
DR   WPI; 90-009296/02.
DR   P-PSDB; R04116, R04117.
PT   Enod2 gene regulatory region - used for early expression of structural
PT   gene in developing root nodule of soybean plant.
PS   Disclosure; page 11-13; 18pp; English.
CC   The sequences contains the Enod2a gene regulatory region (tag a) which is
CC   used to construct a recombinant DNA molecule for the early expression of
CC   structural gene in soybean root nodules. Two open reading frames are
CC   present, ORF1 and ORF2 (R04116 and R04117 resp.) but it is thought that
CC   ORF1 actually encodes the protein, N-75.
SQ   Sequence   3060 BP;   1013 A;   731 C;   406 G;   910 T;
```

```
  Query Match             1.3%;  Score 43.8;  DB 1;  Length 3060;
  Best Local Similarity   51.8%;  Pred. No. 0.0051;
  Matches   99;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy    633 tacatgtctccatacgaccaagctcaaggcccatcaggaccatcatattactcacaacac 692
          ||| |  |||| | ||  || | | |     ||    ||| |  || |   || || |
Db   1917 TACCTACCTCCTCATGAGAAACCGCCACCAGAATACCAACCTCCTCATGAGAAACCACCC 1976

Qy    693 catcaatctccaccacctcatcaccaccatcaccacccgatgccaaaaatccatgagaac 752
          ||| |   |||||||||  | ||||| |  |  ||   || ||||  |   ||  |
Db   1977 CATGAGAATCCACCACCGGAGCACCAACCACCTCATGAGAAGCCACCAGAGCACCAACCA 2036

Qy    753 cctgaacaagtggcatctccatcgattgaagatgctccagagacgaaaccaactcatttg 812
          ||| |  |   | || | |||  |  ||||    |||  ||||    |||| |  | |
Db   2037 CCTCATGAGAAGCCACCACCAGAGTATGAACCACCTCATGAGAAACCACCACCAGAATAC 2096

Qy    813 gttgaaccaca 823
            ||| ||
Db   2097 CAACCACCTCA 2107


RESULT   4
Q03097
ID   Q03097 standard; cDNA; 3856 BP.
AC   Q03097;
DT   05-JUN-1990  (first entry)
DE   Enod2b genomic clone.
KW   Enod2; soybean; root nodule; regulatory region; nodulin 75; ss.
OS   Glycine max (L) Merr. cv. Williams.
FH   Key             Location/Qualifiers
FT   misc_rna        1320. .2365
FT                   /function=regulatory region
FT                   /*tag=  a
FT   cds             2245. .3384
FT                   /*tag=  b
FT                   /label=ORF1
FT                   /note="actual coding sequence of Enod2 encoding N-75"
FT   cds             2552. .3271
FT                   /*tag=  c
FT                   /label=ORF2
PN   EP-349338-A.
PD   03-JAN-1990.
PF   30-JUN-1989; 306666.
PR   01-JUL-1988; US-214297.
PA   (LUBR) Lubrizol Genetics.
PI   Franssen H, Bisseling AH;
DR   WPI; 90-009296/02.
DR   P-PSDB; R04118, R04119.
PT   Enod2 gene regulatory region - used for early expression of structural
PT   gene in developing root nodule of soybean plant.
PS   Disclosure; page 14-16; 18pp; English.
CC   The sequences contains the Enod2a gene regulatory region (tag a) which is
CC   used to construct recombinant DNA molecule for the early expression of a
CC   structural gene in soybean root nodules. Two open reading frames are
CC   present, ORF1 and ORF2 (R04118 and R04119 resp.) but it is thought that
CC   ORF1 actually encodes the protein, N-75.
SQ   Sequence   3856 BP;   1331 A;   757 C;   496 G;   1264 T;   8 Others;


  Query Match             1.3%;  Score 43.8;  DB 1;  Length 3856;
  Best Local Similarity   51.8%;  Pred. No. 0.0059;
  Matches   99;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy    633 tacatgtctccatacgaccaagctcaaggcccatcaggaccatcatattactcacaacac 692
          ||| |  |||| | ||  || | | |     ||    ||| |  || |   || || |
Db   2719 TACCTACCTCCTCATGAGAAACCGCCACCAGAATACCAACCTCCTCATGAGAAACCACCC 2778

Qy    693 catcaatctccaccacctcatcaccaccatcaccacccgatgccaaaaatccatgagaac 752
          ||| |   |||||||||  | ||||| |  |  ||   || ||||  |   ||  |
Db   2779 CATGAGAATCCACCACCGGAGCACCAACCACCTCATGAGAAGCCACCAGAGCACCAACCA 2838

Qy    753 cctgaacaagtggcatctccatcgattgaagatgctccagagacgaaaccaactcatttg 812
          ||| |  |   | || | |||  |  ||||    |||  ||||    |||| |  | |
Db   2839 CCTCATGAGAAGCCACCACCAGAGTATGAACCACCTCATGAGAAACCACCACCAGAATAC 2898

Qy    813 gttgaaccaca 823
            ||| ||
Db   2899 CAACCACCTCA 2909


RESULT   5
Q70140/c
ID   Q70140 standard; cDNA; 672 BP.
AC   Q70140;
DT   23-FEB-1995  (first entry)
DE   Cadrac cDNA.
KW   Cadrac; Cadracpro; Cadop; antisense; Parkinson disease; diagnosis;
KW   therapy; APP gene; amyloid precursor protein; dopamine receptor;
KW   calcium channel; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             1. .672
FT                   /*tag=  a
PN   WO9415967-A.
PD   21-JUL-1994.
PF   30-DEC-1993; E03722.
PR   31-DEC-1992; CA-086576.
PR   06-OCT-1993; US-132169.
PA   (PRED/) PREDDIE R E.
PA   (BERG/) BERGMANN J E.
PI   Bergmann JE,  Preddie RE;
DR   WPI; 94-249143/30.
DR   P-PSDB; R57348.
PT   Agents for prevention and treatment of Parkinson's Disease -
PT   comprises cadrac gene anti-sense to regions of human APP gene
PS   Disclosure; Page 35; 47pp; English.
CC   Cadrac cDNA encodes a protein, Cadop, with potential to be a common
CC   domain for interacting with the human D-2 dopamine receptor and the
CC   human brain Ca2+ channel protein.  Cadrac comprises nucleotides
CC   553-1224 of the antisense strand of the human amyloid precursor
CC   protein mRNA.  Cadrac and Cadop are used for the diagnosis and
CC   therapy of Parkinsonism.
SQ   Sequence   672 BP;    116 A;    210 C;    139 G;    207 T;


  Query Match             1.1%;  Score 38.2;  DB 1;  Length 672;
  Best Local Similarity   61.6%;  Pred. No. 0.1;
  Matches   61;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy    959 ggagaaggcggcgaagattgtgattcaggagacagcggaaggggatgaagatgaggatga 1018
          |||| || |   || ||   |||   |||   ||  ||||  ||||| || ||||| ||
Db    394 GGAGCAGACACAGACTATGCAGATGGTGGAAGAAGAAGAAGCCGATGATGACGAGGACGA 335

Qy   1019 tgaggatagtgattccggggagactatgtctcagggaac 1057
          ||||||| |||||   |  |||     | || ||| | |
Db    334 TGAGGATGGTGATGAGGTAGAGGAAGAGGCTGAGGAACC 296


RESULT   6
T66241
ID   T66241 standard; cDNA; 1981 BP.
AC   T66241;
DT   28-JUL-1997  (first entry)
DE   Romaine lettuce violaxanthin de-epoxidase cDNA.
KW   Violaxanthin de-epoxidase; VDE; light; photosensitivity;
KW   photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
KW   xanthophyll; lettuce; ss.
OS   Lactuca sativa L. cv. romaine.
FH   Key             Location/Qualifiers
FT   misc_difference 26. .29
FT                   /*tag=  a
FT                   /note= "bases 26-29 are illegible in Fig 1"
FT   misc_difference 66. .72
FT                   /*tag=  b
FT                   /note= "bases 66-72 are illegible in Fig 1"
FT   misc_difference 105. .110
```

```
FT                  /*tag=  c
FT                  /note= "bases 105-110 are illegible in Fig 1"
FT   misc_difference 147. .149
FT                  /*tag=  d
FT                  /note= "bases 147-149 are illegible in Fig 1"
FT   misc_difference 186. .189
FT                  /*tag=  e
FT                  /note= "bases 186-189 are illegible in Fig 1"
FT   misc_difference 226. .227
FT                  /*tag=  f
FT                  /note= "bases 226-227 are illegible in Fig 1"
FT   cds             235. .1656
FT                  /*tag=  g
FT   transit_peptide 235. .609
FT                  /*tag=  h
FT   mat_peptide     610. .1653
FT                  /*tag=  i
PN   WO9717447-A2.
PD   15-MAY-1997.
PF   07-NOV-1996; U18291.
PR   07-NOV-1995; US-006315.
PR   06-AUG-1996; US-023502.
PA   (CALJ ) CALGENE INC.
PI   Bugos RC,  Rockholm DC,  Yamamoto HY;
DR   WPI; 97-281036/25.
DR   P-PSDB; W09874.
PT   DNA encoding plant violaxanthin de-epoxidase - used to modify the
PT   sensitivity of a plant to light
PS   Claim 3; Fig 1; 41pp; English.
CC   A cDNA clone (T66241) codes for the 55 kDa violaxanthin de-epoxidase
CC   (VDE) (W09874) of romaine lettuce.  VDE was purified from romaine
CC   lettuce chloroplasts and 2 tryptic peptides were used to develop
CC   primers (see also T66244-45), which amplified a partial VDE
CC   sequence.  The amplified sequence was then used to screen a lettuce
CC   cDNA library, and the 1981 bp DNA sequence was identified.  VDE
CC   nucleic acids (see also T66242-43), in sense or antisense
CC   orientation, can be used in genetic constructs, pref. also contg. a
CC   plastid translocation sequence, to modify VDE levels in plants.
CC   Increased levels result in the plant being tolerant of increased
CC   light and therefore more productive and/or more resistant to
CC   disease.  Underexpression of VDE increases photosynthetic
CC   efficiency under low light.  The photosensitivity of a range of
CC   crops, trees and ornamentals can be modified.
SQ   Sequence   1981 BP;    608 A;    337 C;    433 G;    577 T;
```

```
  Query Match              1.1%;  Score 37.4;  DB 1;  Length 1981;
  Best Local Similarity   62.1%;  Pred. No. 0.38;
  Matches   59;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy     960 gagaaggcggcgaagattgtgattcaggagacagcggaaggggatgaagatgaggatgat 1019
           ||||||  |  ||  |  |    |  |||| ||  |||| || |||| | ||||  ||
Db    1405 GAGAAGTTGTTGATAAAAGAAGCTGTAGAGATAGAAGAAGAGGTTGAAAAAGAGGTGGAG 1464

Qy    1020 gaggatagtgattccggggagactatgtctcaggg 1054
           ||| ||| ||| | | |  |||| ||| |||| |
Db    1465 AAGGTTAGAGATACTGAGATGACTTTGTTTCAGAG 1499
```

```
RESULT   7
T31994
ID   T31994 standard; DNA; 2295 BP.
AC   T31994;
DT   24-NOV-1996  (first entry)
DE   Nonsense-mediated mRNA decay 2 gene C-terminal (1089-3383).
KW   Nonsense-mediated mRNA decay; NMD2; Upf1p; inhibition; ss.
OS   Saccharomyces cerevisiae.
PN   WO9622301-A1.
PD   25-JUL-1996.
PF   27-DEC-1995; U16930.
PR   20-JAN-1995; US-375300.
PA   (UYMA-) UNIV MASSACHUSETTS MEDICAL CENT.
PI   He F,  Jacobson AS;
DR   WPI; 96-354469/35.
DR   P-PSDB; W01897.
PT   Isolated nonsense-mediated mRNA decay gene and protein - used to
PT   develop methods for inhibiting the decay pathway for producing
PT   heterologous or endogenous proteins
PS   Claim 2; Page 52-54; 73pp; English.
CC   The NMD2 gene (T31993) is named after its role in the Nonsense-
CC   Mediated mRNA Decay pathway.  The protein, Nmd2p, binds to Upf1p.
CC   A C-terminal fragment of the protein (T31994) also binds Upf1p and,
CC   when overexpressed in the host cell, the fragment inhibits the
CC   function of Upf1p, thereby inhibiting the nonsense-mediated
CC   mRNA decay pathway.
SQ   Sequence   2295 BP;    891 A;    359 C;    447 G;    598 T;
```

```
  Query Match              1.1%;  Score 37.4;  DB 1;  Length 2295;
  Best Local Similarity   53.0%;  Pred. No. 0.41;
  Matches   80;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy     889 gtgaagttgaacggttaaagaataaaaaatcaggagcatgttcagcgacgaatgggccat 948
           |||  | |  ||    | |  || || |  | |    | || || |||  |  | | ||
Db    1589 GTGTCGATTTACTAGGAGAAGATGAAGACGCGGAGATAAGTACACCGAACACAGAGTCAG 1648

Qy     949 cgaggagtaaggagaaggcggcgaagattgtgattcaggagacagcggaaggggatgaag 1008
           ||    | ||  |  ||||   | |    |  | | | ||   ||  || | ||| || |
Db    1649 CGCCAGGAAAACATCAGGCAAAGCAAGACGAAAGTGAAGATGAAGACGATGAGGACGATG 1708

Qy    1009 atgaggatgatgaggatagtgattccgggga 1039
           | |||||||||| |||   |||  ||  ||
Db    1709 ACGAGGATGATGACGATGACGATGACGATGA 1739
```

```
RESULT   8
T31993
ID   T31993 standard; DNA; 4080 BP.
AC   T31993;
DT   24-NOV-1996  (first entry)
DE   Nonsense-mediated mRNA decay 2 gene.
KW   Nonsense-mediated mRNA decay; NMD2; Upf1p; inhibition; ss.
OS   Saccharomyces cerevisiae.
FH   Key             Location/Qualifiers
FT   cds             362. .3633
FT                   /*tag=  a
FT   intron          368. .480
FT                   /*tag=  b
FT   misc_feature    368. .373
FT                   /*tag=  c
FT                   /note= "5' splice site"
FT   misc_feature    426. .432
FT                   /*tag=  d
FT                   /note= "branchpoint"
FT   misc_feature    478. .480
FT                   /*tag=  e
FT                   /note= "3' splice site"
FT   tata_signal     143. .149
FT                   /*tag=  f
FT   protein_bind    164. .177
FT                   /*tag=  g
FT                   /label= Upf1p_binding_consensus
PN   WO9622301-A1.
PD   25-JUL-1996.
PF   27-DEC-1995; U16930.
PR   20-JAN-1995; US-375300.
PA   (UYMA-) UNIV MASSACHUSETTS MEDICAL CENT.
PI   He F,  Jacobson AS;
DR   WPI; 96-354469/35.
DR   P-PSDB; W04896.
PT   Isolated nonsense-mediated mRNA decay gene and protein - used to
PT   develop methods for inhibiting the decay pathway for producing
PT   heterologous or endogenous proteins
PS   Claim 1; Page 47-49; 73pp; English.
CC   The NMD2 gene (T31993) is named after its role in the Nonsense-
CC   Mediated mRNA Decay pathway.  The protein, Nmd2p, binds to Upf1p.
```

```
CC   A C-terminal fragment of the protein (T31994) also binds Upf1p and,
CC   when overexpressed in the host cell, the fragment inhibits the
CC   function of Upf1p, thereby inhibiting the nonsense-mediated
CC   mRNA decay pathway.
SQ   Sequence   4080 BP;    1449 A;    646 C;    800 G;    1185 T;


  Query Match             1.1%;  Score 37.4;  DB 1;  Length 4080;
  Best Local Similarity   53.0%;  Pred. No. 0.6;
  Matches   80;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy     889 gtgaagttgaacggttaaagaataaaaaatcaggagcatgttcagcgacgaatgggccat 948
           ||| | | ||    | |  || || |  | |    | || || |||  |  | | ||
Db    3038 GTGTCGATTTACTAGGAGAAGATGAAGACGCGGAGATAAGTACACCGAACACAGAGTCAG 3097

Qy     949 cgaggagtaaggagaaggcggcgaagattgtgattcaggagacagcggaaggggatgaaag 1008
           ||    | ||  |  ||||   | |    |  | | | ||   ||  || | ||| || |
Db    3098 CGCCAGGAAAACATCAGGCAAAGCAAGACGAAAGTGAAGATGAAGACGATGAGGACGATG 3157

Qy    1009 atgaggatgatgaggatagtgattccgggga 1039
           | ||||||||||| |||   |||  ||  ||
Db    3158 ACGAGGATGATGACGATGACGATGACGATGA 3188


RESULT   9
X07421
ID   X07421 standard; DNA; 289 BP.
AC   X07421;
DT   08-JUN-1999  (first entry)
DE   Borrelia burgdorferi surface antigen P39.5 clone 12-1 3' end.
KW   Lyme disease; surface antigen; P39.5; diagnosis; prevention;
KW   vaccine; antisense; therapy; treatment; primer; probe;
KW   antibody; DNA; ds.
OS   Borrelia burgdorferi.
PN   WO9900413-A1.
PD   07-JAN-1999.
PF   29-JUN-1998; U13551.
PR   30-JUN-1997; US-051271.
PA   (TULA ) TULANE EDUCATIONAL FUND.
PI   Philipp MT;
DR   WPI; 99-095676/08.
PT   New nucleic acid encoding the P39.5 antigen of Borrelia burgdorferi
PT   - and related vectors, transformants, antibodies and polypeptides,
PT   for diagnosis, prevention and treatment of Lyme disease
PS   Claim 3; Page 59; 89pp; English.
CC   The sequence is that of Borrelia burgdorferi surface antigen P39.5
CC   clone 12-1 (3'end). It can be used in the production of P39.5 or
CC   fragments of it which may be used to raise antibodies to, and in the
CC   development of vaccines against Lyme disease. The sequence can
CC   also be used for making primers and probes for diagnosis, also in
CC   DNA vaccines, as antisense therapeutics and for drug screening.
CC   Antibodies can be used as diagnostic (immunoassay) reagents, for
CC   treating Lyme disease, for affinity purification, for drug
CC   screening and to produce anti-idiotype antibodies (used in the
CC   same way as P39.5 to induce an immune response).
SQ   Sequence   289 BP;    83 A;    29 C;    104 G;    73 T;


  Query Match             1.1%;  Score 36.6;  DB 1;  Length 289;
  Best Local Similarity   55.0%;  Pred. No. 0.19;
  Matches   72;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy     914 aaaatcaggagcatgttcagcgacgaatgggccatcgaggagtaaggagaaggcggcgaa 973
           || | ||  ||   ||   |||| || ||       ||    ||||| ||    || ||
Db     111 AACAGCAAAAGGTGGTGATGCGAAGAGTGTTAATGGGATTGCTAAGGGGATAAAGGGGAT 170

Qy     974 gattgtgattcaggagacagcggaaggggatgaagatgaggatgatgaggatagtgattc 1033
           |||     |   ||||  || || | | | ||||   ||   |||| || | ||||| |
Db     171 TGTTGATGCTGCTGAGAAGGCTGATGCGAAGGAAGGGAAGTTGGATGTGGCTGGTGATGC 230

Qy    1034 cggggagacta 1044
           || || ||||
Db     231 TGGTGAAACTA 241


RESULT  10
Q45606
ID   Q45606 standard; DNA; 1028 BP.
AC   Q45606;
DT   04-DEC-1993  (first entry)
DE   Sequence of pATDFLAG encoding signal peptide, H4VL, linker and
DE   FLAG adapter.
KW   Single chain antibody; SCFV1; human subgroup 4 germline antibody;
KW   variable light; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   cds             293. .775
FT                   /*tag=  a
FT                   /note= "ENCODES SIGNAL,H4VL,LINKER
FT                   see R38320 FT"
FT   cds             784. .816
FT                   /*tag=  c
FT                   /product= flag peptide
PN   WO9312231-A.
PD   24-JUN-1993.
PF   13-DEC-1991; AU0583.
PR   13-DEC-1991; WO-AU0583.
PA   (DOWC ) DOW CHEM AUSTRALIA LTD.
PI   Johnson KS,  Mezes PS,  Richard RA;
DR   WPI; 93-214173/26.
DR   P-PSDB; R38320, R40753.
PT   New composite antibody binding to tumour associated TAG-72
PT   antigen - includes light chain variable region from human
PT   subgroup 4 germline gene, useful, opt. as conjugate, for
PT   diagnosis or treatment of cancer
PS   Example; Figure 28; 150pp; English.
CC   The plasmid pATGFLAG was generated from pSCFVUHH (see Q45605)
CC   to incorporate a flag-coating sequence 3' to any human VH gene
CC   to be expressed contiguously with Hum4 VL. The plasmid pATGFLAG,
CC   when digested with XhoI and Nhe I and purified becomes the human
CC   CH discovery plasmid contg. Hum4 VL in this SCFV format. pSC49FLAG
CC   contains murine antibody CC49VH inserted into sites of XhoI - NheI
CC   of pATDFLAG. It was evaluated for biological activity with the
CC   purpose of serving as a positive control for the FLAG assay
CC   system to detect binding to TAG-72.
SQ   Sequence   1028 BP;    289 A;    226 C;    244 G;    269 T;


  Query Match             1.1%;  Score 36.4;  DB 1;  Length 1028;
  Best Local Similarity   54.7%;  Pred. No. 0.5;
  Matches   93;  Conservative    0;  Mismatches   76;  Indels    1;  Gaps    1;

Qy     969 gcgaagattgtgattcaggagacagcggaaggggatgaagatgaggatgatgaggatagt 1028
           ||| |   || ||   |||| | ||| |  |||||| | | || | || || | ||
Db     704 GCGGACGATGCGAAAAAGGATGCTGCGAAGAAGGATGACGCTAAGAAAGACGATGCTAAA 763

Qy    1029 gattccggggagactatgtctcagggaactactattattgt-tcgaagaccaaaaaccga 1087
           |   |  ||||| ||||| | | ||||| |  |   |   ||  |  |||||||  |
Db     764 AAGGACCTCGAGACAATGTCGCTAGCGACTACAAGGACGATGATGACAAATAAAAACCTA 823

Qy    1088 gcgtcgtacggcacacaatctcatcgaaaagaagtatagatgctcaataa 1137
           |||  | |     || ||  |||||  | | ||    |  || | |  ||
Db     824 GCGATGAATCCGTCAAAACATCATCTTACATAAAGTCACTTGGTGATCAA 873


RESULT  11
X13108/c
ID   X13108 standard; DNA; 11871 BP.
AC   X13108;
DT   19-MAR-1999  (first entry)
DE   Enterococcus faecalis genome contig SEQ ID NO:171.
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
OS   Enterococcus faecalis.
PN   WO9850555-A2.
PD   12-NOV-1998.
```

```
PF   04-MAY-1998; U08985.
PR   14-NOV-1997; US-066009.
PR   06-MAY-1997; US-044031.
PR   16-MAY-1997; US-046655.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Dillon PJ, Kunsch CA;
DR   WPI; 99-045171/04.
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT   - used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
PS   Claim 1; Page 952-958; 2084pp; English.
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   X12938 to X13919 represent these nucleotide sequences which are primary
CC   nucleotide sequences, also known as contigs. The computer-based system
CC   can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
SQ   Sequence   11871 BP;    4127 A;    1887 C;    2503 G;    3347 T;


  Query Match             1.1%;  Score 36.2;  DB 1;  Length 11871;
  Best Local Similarity   47.9%;  Pred. No. 2.9;
  Matches  104;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy    2046 ggtatcactggaaacttcgaagacacctatgaaccatccgccctaacgggcctcttcatg 2105
           |||||||||| ||   |  ||      | | ||  || | |||    |  |  |  | ||
Db   11745 GGTATCACTGAAAGAATTAAATGTTTTTTTCAAGAATTCTCCCACTTGACCACCAACTTC 11686

Qy    2106 tccctctgtgcagtaaaccttgctgaagctgccggagcatcaaacgacggacttccacgc 2165
             |  || ||| |   || | | | ||| || | | |   |  | |     |||     |
Db   11685 GGCAACTTTGCCGCCGACTTGGTTAAAGGTGTCTGTGGCTTGATCTTTCCACTCTTCAGA 11626

Qy    2166 gccgtcatggctcagatctacatttctgcatccatccaatgccgtttggctcttccgaac 2225
           |.  | |      |  | | | ||| | |||||||  |    |  |||    || ||||
Db   11625 GAATTGAGAAGAAACTTTTTCTTTTGTACATCCATTGACCTTCCATTGTTCTTTGCGAAT 11566

Qy    2226 ctactcgcaccattcttctcgggatacttttttacgaa 2262
             |  |       ||||||  | | |  |||||  | |
Db   11565 GTTTTTTAGTTCTTCTTCCAGTGCTGTTTTTTCAGTA 11529


RESULT  12
Q25532
ID   Q25532 standard; DNA; 1648 BP.
AC   Q25532;
DT   22-NOV-1992  (first entry)
DE   Sequence of genomic clone contg. the entire Histidine-rich
DE   protein (HisRP) gene.
KW   Malaria vaccine; Histidine-rich protein; cytoadherence; ss.
OS   Plasmodium lophurae.
FH   Key             Location/Qualifiers
FT   signal_peptide  219. .359
FT                   /*tag=  a
FT   cds             491. .562
FT                   /*tag=  b
FT                   /product= propeptide
FT   intron          360. .490
FT                   /*tag=  c
FT   misc_feature    60. .64
FT                   /*tag=  d
FT                   /label= splice sequence
FT   misc_feature    485. .490
FT                   /*tag=  e
FT                   /label= splice sequence
FT   mat_peptide     563. .1477
FT                   /*tag=  f
PN   US5116965-A.
PD   25-MAY-1992.
PF   26-AUG-1986; 900401.
PR   26-AUG-1986; US-900401.
PA   (SLOK ) SLOAN KETTERING INST CANCER.
PI   Pologe L, Ravetch JV;
DR   WPI; 92-199590/24.
DR   P-PSDB; R24393.
PT   Histidine-rich protein associated with Plasmodium knob phenotype -
PT   and DNA encoding it, used for in vitro diagnosis of P.
PT   Falciparum infection.
PS   Disclosure; Fig 7A-B; 29pp; English.
CC   Two variants of HisRP are produced by P. falciparum.  One is
CC   associated with what is referred to as "knobby phenotype" (K30) and
CC   "knobless phenotype" (K-).  The "knobby" and "knobless" phenotypes
CC   have been implicated in cytoadherence, which is characteristic of
CC   erythrocyte infection.  It has now been found that cDNA expressing
CC   both K+ and K- HisRP can be obtained by the use of P. lophurae HisRP
CC   expressing DNA.  The genomic clone (Q25532) is encoded in two
CC   exons, separating the signal peptide-encoding sequence from the
CC   pro-sequence, confirming that synthesis of the protein occurs via
CC   the preproprotein.  Oligo. probes synthesised to the signal
CC   peptide-encoding exon reveal multiple homologous DNA sequences in
CC   the P. lophurae genome.  The sequence of mature proteins is arranged
CC   in numerous tandem repeats with up to nine histidine residues in a
CC   row, similar to other Plasmodium proteins for which sequence data
CC   have so far been reported.
SQ   Sequence   1648 BP;    569 A;    488 C;    150 G;    441 T;


  Query Match             1.1%;  Score 36;  DB 1;  Length 1648;
  Best Local Similarity   54.8%;  Pred. No. 0.91;
  Matches   92;  Conservative    0;  Mismatches   75;  Indels    1;  Gaps    1;

Qy     561 atgctatcttcaccacaccatacctcaatgacttctcaaccatatacagaagccatggga 620
           | | ||||  || |||      || ||    | || | |||||   ||    |||    |
Db     919 AGGATATCACCATCACCATGCTCCACACCATCATCACCACCATCACCATGCTCCACACCA 978

Qy     621 -catatcaacgggtacatgtctccatacgaccaagctcaaggcccatcaggaccatcata 679
            |||  | ||    ||    ||||| || | ||    ||   ||     |  |||    |
Db     979 TCATCACCACCACCACCATGCTCCACACCATCATCACCACCACCACCATGCTCCACACCA 1038

Qy     680 ttactcacaacaccatcaatctccaccacctcatcaccaccatcacca 727
           | |    || ||||| ||  ||||||  | ||||||||||| |||||
Db    1039 TCATCACCACCACCACCATGCTCCACACCATCATCACCACCACCACCA 1086


RESULT  13
X53491/c
ID   X53491 standard; DNA; 114955 BP.
AC   X53491;
DT   05-JUL-1999  (first entry)
DE   Human adenosine A1 receptor antisense oligonucleotide fragment.
KW   Antisense oligonucleotide; multiple target; antisense treatment;
KW   impaired respiration; inflammation; lung disease;
KW   pulmonary vasoconstriction; inflammation; allergic rhinitis;
KW   acute asthma; allergy; asthma; impeded respiration;
KW   respiratory distress syndrome; pain; cystic fibrosis;
KW   pulmonary hypertension; pulmonary vasoconstriction; emphysema;
KW   chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
KW   colon cancer; breast cancer; lung cancer; pancreatic cancer;
KW   hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW   prostate cancer; ss.
OS   Synthetic.
PN   WO9913886-A1.
PD   25-MAR-1999.
PF   17-SEP-1998; U19419.
PR   09-JUN-1998; US-093972.
PR   17-SEP-1997; US-059160.
PA   (UYEC-) UNIV EAST CAROLINA.
PI   Nyce JW;
```

```
DR   WPI; 99-229400/19.
PT   New antisense oligonucleotides used in treatment of, e.g. pulmonary
PT   vasoconstriction
PS   Disclosure; Page 37; 120pp; English.
CC   The specification describes antisense oligonucleotides (X52869-X55271)
CC   directed against at least 2 mRNAs selected from target genes, coding and
CC   non-coding regions of RNAs corresponding to target genes, gene
CC   initiation codons, genomic flanking regions, intron-exon borders, the
CC   5'-end, the 3'-end and the juxta-section between coding and non-coding
CC   regions and all segments of RNAs encoding proteins associated with one
CC   or more diseases, conditions or mixtures. The antisense oligonucleotides
CC   may be derived from sequences X55272-74. These multiple target
CC   oligonucleotides (specifically X55180-271) can be used for the antisense
CC   treatment of diseases and conditions. Typical diseases and conditions
CC   are those associated with impaired respiration and inflammation,
CC   including lung diseases, pulmonary vasoconstriction, inflammation,
CC   allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,
CC   respiratory distress syndrome, pain, cystic fibrosis, pulmonary
CC   hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
CC   carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic
CC   cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic
CC   metastases, as well as all types of cancers which may metastasize or have
CC   metastasized to the lungs, including breast and prostate cancer.
SQ   Sequence   114955 BP;    6071 A;    29417 C;    36712 G;    21328 T;


  Query Match             1.0%;  Score 35.6;  DB 1;  Length 114955;
  Best Local Similarity   38.0%;  Pred. No. 20;
  Matches   89;  Conservative   34;  Mismatches  111;  Indels    0;  Gaps    0;

Qy    687 caacaccatcaatctccaccacctcatcaccaccatcaccacccgatgccaaaaatccat 746
          |: |:| |   :   :| :  ||  : |::|| ::  |:||  | | : |: ||| : |
Db  79264 CVGCVCAAGNNHNNNSCCVGGCCGGVGCVVCAGVVAGAVCAAACCAVVGCVGAAACVGAA 79205

Qy    747 gagaaccctgaacaagtggcatctccatcgattgaagatgctccagagacgaaaccaact 806
          ||| || : : | |:  : ||   |||  ||  :  || ::   :||: : ||::: ||
Db  79204 GAGGACAVGVCAAAVAVVACAGAVCCACAGAVGVGGGAVVVVGAVGAVCVAAAVVVCACV 79145

Qy    807 catttggttgaaccacaaagtccaaaaagcccgcagaatatgaaagaggagcttcttcgg 866
              |     : || |:     ::| ||||| : :  :        |  | | |||| |
Db  79144 GGCAVGCCACCVGCAGAVGAAGAVVACAGCCCCVGVNNHNNNSCCTGGCCGGTGCTTCAG 79085

Qy    867 ttactagttaacatgtctccgagtgaagttgaacggttaaagaataaaaaatca 920
          |||      | |||  ||   | |||||  ||   || ||| | || | |  ||
Db  79084 TTAGATCAAACCATTGCTGAAACTGAAGAGGACATGTCAAATATTACAGATCCA 79031


RESULT  14
T11575/c
ID   T11575 standard; cDNA; 2758 BP.
AC   T11575;
DT   09-AUG-1996  (first entry)
DE   Human glucagon degrading enzyme gene.
KW   Glucagon degrading enzyme; catalyst; cleavage; selectin; human; primer;
KW   vasoactive intestinal peptide; VIP; pancreatic carcimoma cell line; PCR;
KW   amplification; polymerase chain reaction; probe; expression vector;
KW   eukaryote; SV40 promoter; COS-7; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             100. .2694
FT                   /*tag=  a
FT                   /product= glucagon degrading enzyme
PN   J08023972-A.
PD   30-JAN-1996.
PF   19-JUL-1994; 187936.
PR   19-JUL-1994; JP-187936.
PA   (SUNR ) SUNTORY LTD.
DR   WPI; 96-133414/14.
DR   P-PSDB; R93021.
PT   New glucagon decomposing enzyme, and DNA encoding it - for
PT   specifically cleaving glucagon and vasoactive intestinal peptide, in
PT   the prevention and treatment of diseases caused by excess glucagon
PT   and VIP
PS   Claim 4; Page 3; 18pp; Japanese.
CC   This is the nucleotide sequence of a novel isolated gene encoding a
CC   glucagon degrading enzyme (GDE) of mol. wt. 83 kD.  The enzyme has a
CC   pH optimum of 6.8 and catalyses the cleavage of glucagon, vasoactive
CC   intestinal peptide and selectin (R93022-4).  The gene was isolated
CC   from a human pancreatic carcimoma cell line HPC-Yo cDNA library by
CC   screening the library with an anti-GDE peptide antibody, amplifying
CC   the inserts with the primers T18903-4 and probing the fragments with
CC   the probe T18905.  This screening resulted in the full length clone
CC   designated lambda GDE4-2.  The coding region of the clone was
CC   subsequently PCR amplified by the primers T11576-7 and inserted into
CC   the eukaryotic expression vector pKDCR under control of the SV40
CC   promoter for production of the protein in COS-7 cells.  The protein
CC   is useful in preventing and treating diseases characterised by an
CC   excess of glucagon or vasoactive intestinal peptide.
SQ   Sequence   2758 BP;    968 A;    416 C;    483 G;    891 T;


  Query Match             1.0%;  Score 34.8;  DB 1;  Length 2758;
  Best Local Similarity   65.4%;  Pred. No. 3;
  Matches   51;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy   1650 tatgtctgggtgttcaacatcttaatgatcatatatgtggttgtcaaactgctgatccat 1709
          | | |||     |||||||||||| ||  | ||||| ||     |||||||  ||||  ||
Db    696 TCTATCTTCTAATTCAACATCTTTATTTTGATATAAGTCTACATCAAACTCATGATTAAT 637

Qy   1710 ggtgaccctgttcaagac 1727
           || || | ||| || ||
Db    636 TGTCACGCCGTTTAAAAC 619


RESULT  15
X19798
ID   X19798 standard; cDNA; 2344 BP.
AC   X19798;
DT   09-JUN-1999  (first entry)
DE   Human growth regulator protein GRREG encoding cDNA.
KW   Human; growth regulator protein; GRREG; cancer; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             226. .2250
FT                   /*tag=  a
PN   WO9902680-A1.
PD   21-JAN-1999.
PF   30-JUN-1998; U13409.
PR   11-JUL-1997; US-893852.
PA   (INCY-) INCYTE PHARM INC.
PI   Bandman O, Corley NC, Lal P, Shah P;
DR   WPI; 99-120881/10.
DR   P-PSDB; W99891.
PT   Novel human growth regulator protein - useful in the treatment of
PT   cancers
PS   Claim 7; Fig 1; 70pp; English.
CC   The present sequence encodes a human growth regulator protein,
CC   designated GRREG. The protein can be used in methods to treat cancer.
CC   The protein can also be used in binding assays to detect (ant)agonists
CC   of GRREG. A fragment from the nucleic acid sequence encoding GRREG can
CC   be used as a probe for detecting GRREG encoding sequences (especially
CC   in PCR amplified samples).
SQ   Sequence   2344 BP;    556 A;    667 C;    709 G;    412 T;


  Query Match             1.0%;  Score 34.4;  DB 1;  Length 2344;
  Best Local Similarity   78.8%;  Pred. No. 3.6;
  Matches   41;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy    985 aggagacagcggaaggggatgaagatgaggatgatgaggatagtgattccgg 1036
          |||| |||| ||||| ||| || || || || ||||||| ||||| || ||
Db   1268 AGGACACAGAGGAAGAGGAAGATGAGGAAGAAGATGAGGACAGTGACTCTGG 1319
```

Search completed: September 4, 2000, 00:38:25
Job time: 16880 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2000, 00:24:33 ; Search time 4457.97 Seconds
(without alignments)
1949.715 Million cell updates/sec

Title: US-09-332-522B-3
Perfect score: 1971
Sequence: 1 cggcacgaggattaatgctg.....aaaaaaaaaaaaaaaaaaac 1971

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: em_est35:*
83: em_est36:*
84: em_est37:*
85: gb_est48:*
86: gb_est49:*
87: gb_est50:*
88: gb_est51:*
89: gb_est52:*
90: gb_est53:*
91: gb_est54:*
92: gb_est55:*
93: gb_gss1:*
94: gb_gss2:*
95: gb_gss3:*
96: gb_gss4:*
97: em_gss1:*
98: em_gss2:*
99: em_gss3:*
100: em_gss4:*
101: gb_gss5:*
102: gb_gss6:*
103: gb_gss7:*
104: gb_gss8:*
105: gb_gss9:*
106: em_gss5:*
107: em_gss6:*
108: em_gss7:*
109: em_gss8:*
110: em_gss9:*
111: em_gss10:*
112: em_gss11:*
113: gb_gss10:*
114: gb_gss11:*
115: em_gss12:*
116: gb_gss12:*

117: gb_gss13:*
118: gb_gss14:*
119: gb_gss15:*
120: gb_gss16:*
121: gb_gss17:*
122: gb_gss18:*
123: gb_gss19:*
124: em_gss13:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	745.4	37.8	788	25	AA391707	AA391707 LD11632.5
2	590.4	30.0	592	41	AI546088	AI546088 LD39580.5
3	564.2	28.6	578	26	AA439767	AA439767 LD14421.5
4	557.6	28.3	584	33	AA990888	AA990888 LD34294.5
5	50.6	2.6	391	45	AI852939	AI852939 UI-M-BH0-
c 6	46.8	2.4	1101	122	CNS00Z93	AL097377 Drosophil
7	46.4	2.4	928	122	CNS00DKY	AL071865 Drosophil
8	46.2	2.3	282	81	C94400	C94400 C94400 Dict
c 9	45.8	2.3	946	122	CNS008EV	AL052494 Drosophil
10	45.4	2.3	455	80	C24663	C24663 C24663 Dict
c 11	45.2	2.3	928	122	CNS00DKY	AL071865 Drosophil
12	44.6	2.3	311	48	AU071134	AU071134 AU071134
c 13	44.6	2.3	825	122	CNS00BTL	AL058023 Drosophil
14	44.4	2.3	315	31	AA839308	AA839308 vo40d03.r
15	43.8	2.2	202	70	AW307211	AW307211 sf54b11.y
16	43.8	2.2	360	80	C11831	C11831 C11831 Yuji
17	43.8	2.2	437	80	AW702259	AW702259 TgESTzz74
c 18	43.8	2.2	995	122	CNS003IJ	AL064551 Drosophil
c 19	43.6	2.2	427	45	AI874339	AI874339 tz64e05.x
c 20	43.4	2.2	240	48	AU073211	AU073211 AU073211
21	43.2	2.2	1201	123	CNS016A0	AL106482 Drosophil
c 22	42.6	2.2	522	42	AI673293	AI673293 tw69g09.x
c 23	42.2	2.1	348	40	AI496844	AI496844 fb64e06.x
24	42.2	2.1	563	48	AU038220	AU038220 AU038220
25	42.2	2.1	726	39	AI406223	AI406223 GH26449.5
c 26	42	2.1	213	37	AI251434	AI251434 qu74e10.x
27	42	2.1	264	81	C90816	C90816 C90816 Dict
c 28	42	2.1	312	46	AI928293	AI928293 wp12a09.x
29	42	2.1	517	47	AU033605	AU033605 AU033605
c 30	42	2.1	786	122	CNS006O9	AL065615 Drosophil
c 31	41.8	2.1	197	41	AI559457	AI559457 tq44b03.x
32	41.8	2.1	326	23	AA242229	AA242229 mw26f02.r
33	41.6	2.1	245	70	AW311640	AW311640 sg43d06.y
c 34	41.4	2.1	507	64	AW165325	AW165325 fd99f02.y
35	41.4	2.1	690	81	C93837	C93837 C93837 Dict
36	41.4	2.1	1101	122	CNS003BD	AL064091 Drosophil
37	41.2	2.1	200	81	C93770	C93770 C93770 Dict
c 38	41.2	2.1	312	31	AA780779	AA780779 ag14b08.s
39	41.2	2.1	314	63	AW101789	AW101789 sd70d10.y
c 40	41.2	2.1	351	23	AA253430	AA253430 zr77f02.s
c 41	41.2	2.1	761	44	AI812108	AI812108 tw77d04.x
42	41.2	2.1	908	123	CNS01678	AL106382 Drosophil
c 43	41	2.1	286	39	AI416651	AI416651 sa10a03.x
44	41	2.1	301	47	AL121328	AL121328 DKFZp762P
45	41	2.1	467	31	AA801779	AA801779 GM12988.5

ALIGNMENTS

```
RESULT  1
AA391707
LOCUS       AA391707      788 bp    mRNA            EST       28-NOV-1998
DEFINITION  LD11632.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD11632 5prime, mRNA sequence.
ACCESSION   AA391707
VERSION     AA391707.1  GI:2044683
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 788)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797339.
            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 116  row: C  column: 8
            High quality sequence stop: 694.
FEATURES             Location/Qualifiers
     source          1. .788
                     /organism="Drosophila melanogaster"
                     /db_xref="BDGP_EST:BDcln010843"
                     /db_xref="taxon:7227"
                     /clone="LD11632"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                     EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      175 a    223 c    190 g    200 t
ORIGIN
```

Query Match 37.8%; Score 745.4; DB 25; Length 788;
Best Local Similarity 98.9%; Pred. No. 1.4e-202;
Matches 782; Conservative 0; Mismatches 6; Indels 3; Gaps 3;

```
Qy      19 tgatttctggtctggactacacagcattgctggtataaggagtcgggaccagaggagtaa 78
           ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       1 TGATTTCTGGTCTGGACTACACAGCATTGC-GGTATAAGGAGTCGGGACCAGAGGAGTAA 59

Qy      79 gatttcgggaaggaatcccgtccggtagggactactagcattcgcaagtgacgtccagca 138
           ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db      60 GATTTCGGGAAGGAATCCCGTCCGGTA-GGACTACTAGCATTCGCAAGTGACGTCCAGCA 118

Qy     139 accggaggaccccccaactgtagaatccgcatcaccatcctaatcccaacaaaccaatgac 198
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     119 ACCGGA-GACCCCCAACTGTAGAATCCGCATCACCATCCTAATCCCAACAAACCAATGAC 177

Qy     199 atcttgagacctcaccagccatggatcccttcgtgttcttcatagtactggcatcgcttt 258
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     178 ATCTTGAGACCTCACCAGCCATGGATCCCTTCGTGTTCTTCATAGTACTGGCATCGCTTT 237

Qy     259 atggcgttctttacttttcgaccgcttcttcaagagttgcatgcactacccgtacgatg 318
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     238 ATGGCGTTCTTTACTTTTCGACCGCTTCTTCAAGAGTTGCATGCACTACCCGTACGATG 297

Qy     319 ccttcctcaagaacaccgggctgagtataaatttcatgagcctccactggcacacgagtg 378
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     298 CCTTCCTCAAGAACACCGGGCTGAGTATAAATTTCATGAGCCTCCACTGGCACACGAGTG 357

Qy     379 cctttaacaggaccctcctacgctggggatctgccggtaacagctgcacccggagagtaa 438
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     358 CCTTTAACAGGACCCTCCTACGCTGGGGATCTGCCGGTAACAGCTGCACCCGGAGAGTAA 417

Qy     439 tgatcaccagctttaatgtaggagtcctggtcacctttctctgctcccgatcggtctga 498
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    418 TGATCACCAGCTTTAATGTAGGAGTCCTGGTCACCTTTTCTCTGCTCCCGATCGGTCTGA 477

Qy    499 tcctgctcattgccactatcttcagcagtggtgaacaagacagctcttcgtctgtatcct 558
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    478 TCCTGCTCATTGCCACTATCTTCAGCAGTGGTGAACAAGACAGCTCTTCGTCTGTATCCT 537

Qy    559 cgcccgttggagtccctgtgcagctggaaattctactgcccggcgtcaacttgccgttgg 618
          ||||||||||||   ||||||||||||||||||||||| ||||||||||||||||||||
Db    538 CGCCCGTTGGATGTCCTGTGCAGCTGGAAATTCTACTGCTCGGCGTCAACTTGCCGTTGG 597

Qy    619 aggagatcggatactacatcacaacccttgtgctctgcttggtggtgcacgagatgggac 678
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    598 AGGAGATCGGATACTACATCACAACCCTTGTGCTCTGCTTGGTGGTGCACGAGATGGGAC 657

Qy    679 acgccctggccgctgtgatggaggatgtgcctgtcaccgggtttggaataaagttcatct 738
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    658 ACGCCCTGGCCGCTGTGATGGAGGATGTGCCTGTCACCGGGTTTGGAATAAAGTTCATCT 717

Qy    739 tctgcctgccgttagcatacacggagctctcccacgaccacttaaacagtctacgttggt 798
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    718 TCTGCCTGCCGTTAGCATACACGGAGCTCTCCCACGACCACTTAAACAGTCTACGGTGGG 777

Qy    799 tccgcaagcta 809
          ||||||||||||
Db    778 TCCGCAAGCTA 788


RESULT   2
AI546088
LOCUS       AI546088      592 bp    mRNA            EST       22-MAR-1999
DEFINITION  LD39580.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD39580 5prime, mRNA sequence.
ACCESSION   AI546088
VERSION     AI546088.1  GI:4463461
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 592)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic sequence AC005465; hit genomic sequence AC005466
            Plate: 395  row: G  column: 8
            High quality sequence stop: 511.
FEATURES             Location/Qualifiers
     source          1. .592
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LD39580"
                     /clone_lib="LD Drosophila melanogaster embryo pOT2"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="XL1 Blue"
                     /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2. "
BASE COUNT      142 a    150 c    159 g    141 t
ORIGIN


  Query Match             30.0%;  Score 590.4;  DB 41;  Length 592;
  Best Local Similarity   99.8%;  Pred. No. 3.2e-158;
  Matches  591;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    597 cccggcgtcaacttgccgttggaggagatcggatactacatcacaacccttgtgctctgc 656
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CCGGGCGTCAACTTGCCGTTGGAGGAGATCGGATACTACATCACAACCCTTGTGCTCTGC 60

Qy    657 ttggtggtgcacgagatgggacacgccctggccgctgtgatggaggatgtgcctgtcacc 716
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TTGGTGGTGCACGAGATGGGACACGCCCTGGCCGCTGTGATGGAGGATGTGCCTGTCACC 120

Qy    717 gggtttggaataaagttcatcttctgcctgccgttagcatacacggagctctcccacgac 776
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGGTTTGGAATAAAGTTCATCTTCTGCCTGCCGTTAGCATACACGGAGCTCTCCCACGAC 180

Qy    777 cacttaaacagtctacgttggttccgcaagctacgtgttctgtgcgctggaatctggcat 836
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CACTTAAACAGTCTACGTTGGTTCCGCAAGCTACGTGTTCTGTGCGCTGGAATCTGGCAT 240

Qy    837 aattttgtgttcgctggcgtgtgctatctcttaatctcaacggtgggaatcactatgtca 896
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AATTTTGTGTTCGCTGGCGTGTGCTATCTCTTAATCTCAACGGTGGGAATCACTATGTCA 300

Qy    897 cctttgtacgcttacaaccaacacgtagtggtcactgaactaacaaggaaatccccgctg 956
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCTTTGTACGCTTACAACCAACACGTAGTGGTCACTGAACTAACAAGGAAATCCCCGCTG 360

Qy    957 aggggagagcgcggcttgcaagtggacaatcaaataacccaagtaaacggctgcccagta 1016
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AGGGGAGAGCGCGGCTTGCAAGTGGACAATCAAATAACCCAAGTAAACGGCTGCCCAGTA 420

Qy   1017 aacagcgaggagagttgggtgacatgcctgcagaactctctgaagctcaagccgggctac 1076
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AACAGCGAGGAGAGTTGGGTGACATGCCTGCAGAACTCTCTGAAGCTCAAGCCGGGCTAC 480

Qy   1077 tgtgtgagtgcggacttcgtgcagcttaacgacgaaagcagcgccatctcacatcatagc 1136
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TGTGTGAGTGCGGACTTCGTGCAGCTTAACGACGAAAGCAGCGCCATCTCACATCATAGC 540

Qy   1137 attgatggtcagctacagtgctgtgatgaactaaatccgaacgtaagctgct 1188
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ATTGATGGTCAGCTACAGTGCTGTGATGAACTAAATCCGAACGTAAGCTGCT 592


RESULT   3
AA439767
LOCUS       AA439767      578 bp    mRNA            EST       28-NOV-1998
DEFINITION  LD14421.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD14421 5prime, mRNA sequence.
ACCESSION   AA439767
VERSION     AA439767.1  GI:2151648
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 578)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 144  row: B  column: 9
            High quality sequence stop: 434.
FEATURES             Location/Qualifiers
     source          1. .578
                     /organism="Drosophila melanogaster"
                     /db_xref="BDGP_EST:BDcln013616"
```

```
                     /db_xref="taxon:7227"
                     /clone="LD14421"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                     EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      134 a    163 c    135 g    146 t
ORIGIN


  Query Match             28.6%;  Score 564.2;  DB 26;  Length 578;
  Best Local Similarity   98.6%;  Pred. No. 1e-150;
  Matches  569;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      11 attaatgctgatttctggtctggactacacagcattgctggtataaggagtcgggaccag 70
           || |  || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 ATAATGGCGGATTTCTGGTCTGGACTACACAGCATTGCTGGTATAAGGAGTCGGGACCAG 61

Qy      71 aggagtaagatttcgggaaggaatcccgtccggtagggactactagcattcgcaagtgac 130
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 AGGAGTAAGATTTCGGGAAGGAATCCCGTCCGGTAGGGACTACTAGCATTCGCAAGTGAC 121

Qy     131 gtccagcaaccggaggacccccaactgtagaatccgcatcaccatcctaatcccaacaaa 190
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 GTCCAGCAACCGGAGGACCCCCAACTGTAGAATCCGCATCACCATCCTAATCCCAACAAA 181

Qy     191 ccaatgacatcttgagacctcaccagccatggatcccttcgtgttcttcatagtactggc 250
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 CCAATGACATCTTGAGACCTCACCAGCCATGGATCCCTTCGTGTTCTTCATAGTACTGGC 241

Qy     251 atcgctttatggcgttctttacttttttcgaccgcttcttcaagagttgcatgcactaccc 310
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 ATCGCTTTATGGCGTTCTTTACTTTTTCGACCGCTTCTTCAAGAGTTGCATGCACTACCC 301

Qy     311 gtacgatgccttcctcaagaacaccgggctgagtataaatttcatgagcctccactggca 370
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 GTACGATGCCTTCCTCAAGAACACCGGGCTGAGTATAAATTTCATGAGCCTCCACTGGCA 361

Qy     371 cacgagtgcctttaacaggaccctcctacgctggggatctgccggtaacagctgcacccg 430
           |||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     362 CACGAGTGCTCTTAACAGGACCCTCCTACGCTGGGGATCTGCCGGTAACAGCTGCACCCG 421

Qy     431 gagagtaatgatcaccagctttaatgtaggagtcctggtcacctttttctctgctcccgat 490
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     422 GAGAGTAATGATCACCAGCTTTAATGTAGGAGTCCTGGTCACCTTTTCTCTGCTCCCGAT 481

Qy     491 cggtctgatcctgctcattgccactatcttcagcagtggtgaacaagacagctcttcgtc 550
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 CGGTCTGATCCTGCTCATTGCCACTATCTTCAGCAGTGGTGAACAAGACAGCTCTTCGTC 541

Qy     551 tgtatcctcgcccgttggagtccctgtgcagctggaa 587
           ||||| |||||||||||||||||||||||||||||| |
Db     542 TGTATACTCGCCCGTTGGAGTCCCTGTGCAGCTGGGA 578


RESULT   4
AA990888
LOCUS       AA990888      584 bp    mRNA            EST       24-NOV-1998
DEFINITION  LD34294.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD34294 5prime, mRNA sequence.
ACCESSION   AA990888
VERSION     AA990888.1  GI:3177138
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 584)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Apr 7, 1998 this sequence version replaced gi:3036104.
            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 342  row: H  column: 10
            High quality sequence stop: 583.
FEATURES             Location/Qualifiers
     source          1. .584
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LD34294"
                     /clone_lib="LD Drosophila melanogaster embryo pOT2"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="XL1 Blue"
                     /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2. "
BASE COUNT      137 a    164 c    130 g    153 t
ORIGIN


  Query Match             28.3%;  Score 557.6;  DB 33;  Length 584;
  Best Local Similarity   98.3%;  Pred. No. 7.9e-149;
  Matches  574;  Conservative    0;  Mismatches    9;  Indels    1;  Gaps    1;

Qy      11 attaatgctgatttctggtctggactacacagcattgctggtataaggagtcgggaccag 70
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATTAATGCTGATTTCTGGTCTGGACTACACAGCATTGCTGGTATAAGGAGTCCGGACCAG 60

Qy      71 aggagtaagatttcgggaaggaatcccgtccggt-agggactactagcattcgcaagtga 129
           | |||||||||||| || | | |||||||||||| ||||| || ||| ||||||||||||
Db      61 AAGAGTAAGATTTCCGGGAAGGATCCCGTCCGGTAAGGGATTATTAGAATTCGCAAGTGA 120

Qy     130 cgtccagcaaccggaggacccccaactgtagaatccgcatcaccatcctaatcccaacaa 189
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CGTCCAGCAACCGGAGGACCCCCAACTGTAGAATCCGCATCACCATCCTAATCCCAACAA 180

Qy     190 accaatgacatcttgagacctcaccagccatggatcccttcgtgttcttcatagtactgg 249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ACCAATGACATCTTGAGACCTCACCAGCCATGGATCCCTTCGTGTTCTTCATAGTACTGG 240

Qy     250 catcgctttatggcgttctttacttttttcgaccgcttcttcaagagttgcatgcactacc 309
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CATCGCTTTATGGCGTTCTTTACTTTTTCGACCGCTTCTTCAAGAGTTGCATGCACTACC 300

Qy     310 cgtacgatgccttcctcaagaacaccgggctgagtataaatttcatgagcctccactggc 369
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CGTACGATGCCTTCCTCAAGAACACCGGGCTGAGTATAAATTTCATGAGCCTCCACTGGC 360

Qy     370 acacgagtgcctttaacaggaccctcctacgctggggatctgccggtaacagctgcaccc 429
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ACACGAGTGCCTTTAACAGGACCCTCCTACGCTGGGGATCTGCCGGTAACAGCTGCACCC 420

Qy     430 ggagagtaatgatcaccagctttaatgtaggagtcctggtcacctttttctctgctcccga 489
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GGAGAGTAATGATCACCAGCTTTAATGTAGGAGTCCTGGTCACCTTTTCTCTGCTCCCGA 480

Qy     490 tcggtctgatcctgctcattgccactatcttcagcagtggtgaacaagacagctcttcgt 549
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TCGGTCTGATCCTGCTCATTGCCACTATCTTCAGCAGTGGTGAACAAGACAGCTCTTCGT 540

Qy     550 ctgtatcctcgcccgttggagtccctgtgcagctggaaattcta 593
           |||||||||||||||||||||||||||||||||||||||||||||
Db     541 CTGTATCCTCGCCCGTTGGAGTCCCTGTGCAGCTGGAAATTCTA 584
```

```
RESULT   5
AI852939
LOCUS       AI852939      391 bp    mRNA            EST       15-JUL-1999
DEFINITION  UI-M-BH0-aiz-d-03-0-UI.s1 NIH_BMAP_M_S1 Mus musculus cDNA clone
            UI-M-BH0-aiz-d-03-0-UI 3', mRNA sequence.
ACCESSION   AI852939
VERSION     AI852939.1  GI:5496845
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 391)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     On May 7, 1998 this sequence version replaced gi:3118977.
            Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized corpus striatum library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
            available by the means that is soon to be determined. When NIH
            determines the means for distribution of the BMAP cDNA clones, this
            record will be updated accordingly when that means is determined.
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .391
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH0-aiz-d-03-0-UI"
                     /clone_lib="NIH_BMAP_M_S1"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_M_S1 library is a subtracted library derived from
                     a mixture of normalized libraries from ten regions of the
                     mouse brain (cerebellum, brain stems, olfactory bulbs,
                     hypothalamus, cortex, amygdala, basal ganglia, pineal
                     gland, striatum, hipoccampus). The driver used for
                     subtraction consisted of a pool of 20,000 cDNA clones
                     obtained from non-normalized and normalized libraries of
                     these ten regions of the mouse brain.;
                     TAG_LIB=NIH_BMAP_M_S1; TAG_TISSUE=corpus-striatum;
                     TAG_SEQ=ACGGC"
BASE COUNT       88 a     88 c     76 g    139 t
ORIGIN


  Query Match             2.6%;  Score 50.6;  DB 45;  Length 391;
  Best Local Similarity   47.4%;  Pred. No. 0.0012;
  Matches  152;  Conservative    0;  Mismatches  169;  Indels    0;  Gaps    0;

Qy    732 ttcatcttctgcctgccgttagcatacacggagctctcccacgaccacttaaacagtcta 791
          ||  | || |     ||    ||||     || || | |  |  ||| ||  |   | ||
Db      3 TTTTTTTTTTTTTATCCTGGGGCATTTGTTGATCTGTTCACCACCCATTTGCAACTTATA 62

Qy    792 cgttggttccgcaagctacgtgttctgtgcgctggaatctggcataattttgtgttcgct 851
                ||| |||||  |  |  | || || || || ||||||||||||||  | |||
Db     63 TCACCAGTCCAGCAGCTAAGGATATTTTGTGCAGGTATATGGCATAATTTTGTCCTTGCT 122

Qy    852 ggcgtgtgctatctcttaatctcaacggtgggaatcactatgtcacctttgtacgcttac 911
            | |  |   | |      |         |   | | |   | |  ||||| |||
Db    123 CTCCTTGGTATTTTAGCTCTTGTTCTCCTCCCAGTAATCCTCTTGCCTTTTTACTACACT 182

Qy    912 aaccaacacgtagtggtcactgaactaacaaggaaatccccgctgagggagagcgcggc 971
                |||  |  |||||||| |  ||  | | || ||    |  |||    | |||
Db    183 GGAGTTGGGGTACTTATCACTGAAGTCGCAGAGGACTCACCTGCCATTGGACCCAGAGGC 242

Qy    972 ttgcaagtggacaatcaaataacccaagtaaacggctgcccagtaaacagcgaggagagt 1031
            |    ||||   | |   | || ||  || | |  || || || |  |  | | |   |
Db    243 CTTTTGTGGGAGACCTTGTCACTCATCTACAGGATTGTCCCGTTACTAATGTGCAAGAT 302

Qy   1032 tgggtgacatgcctgcagaac 1052
          |||     |||  |  | | |
Db    303 TGGAATGAATGTTTAGACACC 323


RESULT   6
CNS00Z93/c
LOCUS       CNS00Z93     1101 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN01G23 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL097377
VERSION     AL097377.1  GI:5608988
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN01G23"
                     /note="end : T7"
BASE COUNT      441 a    129 c    135 g    274 t    122 others
ORIGIN


  Query Match             2.4%;  Score 46.8;  DB 122;  Length 1101;
  Best Local Similarity   43.6%;  Pred. No. 0.02;
  Matches   61;  Conservative   28;  Mismatches   51;  Indels    0;  Gaps    0;

Qy   1831 acaaaaacacagcgaattgtagcacctcaaagattcgatagctttttgtcatagtcctta 1890
          | |||||  | :   ||:| |   :  |  :   ||  :::  |::|| :  |  |  ||:
Db   1092 AWAAAAAAAAAWATAAAWTATTTTWTATTTWTTKTTTMWWWTWTWWTTTYTTTTTTTTTTW 1033

Qy   1891 gtcttaactcgtatttatttttcgtacggttgtcgagctcaaaaataaaatcaaattaagc 1950
            : || : :  ||||| |||:  |    || |  |  : ||||::|:::::| |::|:
Db   1032 TWWTTTWTWAATATTTTTTTWTTTYTTTTTTTTAATWWAAAAAWWAWWWWVATAWWAWAA 973

Qy   1951 taaaaaaaaaaaaaaaaaaa 1970
```

```
          |||:|||||||||||||:|
Db    972 AAAAWAAAAAAAAAAAAWA 953


RESULT   7
CNS00DKY
LOCUS       CNS00DKY      928 bp    DNA             GSS       04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL071865
VERSION     AL071865.1  GI:4948170
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 928)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .928
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR27A24"
                     /note="end : T7"
BASE COUNT      262 a     70 c     84 g    321 t    191 others
ORIGIN


  Query Match             2.4%;  Score 46.4;  DB 122;  Length 928;
  Best Local Similarity   36.9%;  Pred. No. 0.025;
  Matches   59;  Conservative   38;  Mismatches   63;  Indels    0;  Gaps    0;

Qy   1811 atttcaccttcacgaaacacacaaaaacacagcgaattgtagcacctcaaagattcgata 1870
           :|:| |  |  |  :|| | | ||||: | |   |||| ||  |    ||: |     ||
Db    647 WTWTTATATAWAAAWAAAAWAAAAAAWAATAAAAAATTTTAWAAWTAAAAWAAAAATTTA 706

Qy   1871 gctttttgtcatagtccttagtcttaactcgtatttatttttcgtacggttgtcgagctca 1930
          |||||  |  |  |   |: : :  |    |||:::::   |: |: :    :  : :
Db    707 WTTTTTTTTTTTTTTWTATWTWAWATAWAAATATWWWWWWATWTWDGKNWNNNAWWWWWW 766

Qy   1931 aaaataaaatcaaattaagctaaaaaaaaaaaaaaaaaa 1970
          :::: ::::  |::::::  :::||||||||||||||:|||
Db    767 WWWWAWWWWAAAWWWWWWWWWWAAAAAAAAAAAAAAWAAA 806


RESULT   8
C94400
LOCUS       C94400        282 bp    mRNA            EST       15-JUN-1998
DEFINITION  C94400 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSK649, mRNA sequence.
ACCESSION   C94400
VERSION     C94400.1  GI:3219015
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 282)
  AUTHORS   Yoshino,R., Morio,T. and Tanaka,Y.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 6, 2000 this sequence version replaced gi:6677473.
            Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
            Email: d402hu@sakura.cc.tsukuba.ac.jp
            PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES             Location/Qualifiers
     source          1. .282
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SSK649"
                     /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
                     /dev_stage="slug"
BASE COUNT      151 a     24 c     19 g     87 t      1 others
ORIGIN


  Query Match             2.3%;  Score 46.2;  DB 81;  Length 282;
  Best Local Similarity   53.3%;  Pred. No. 0.02;
  Matches   96;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy   1791 tgcaaagacgctagactgctatttcaccttcacgaaacacacaaaaacacagcgaattgt 1850
           | ||| |||   |   ||| | |  |  || |  ||  | | ||||| | |  | ||| |
Db     83 TTCAACGACTTCAAGTTGCNAATAAATGTTTAATAATAATAAAAAAAAAAAATGTATTTT 142

Qy   1851 agcacctcaaagattcgatagctttttgtcatagtccttagtcttaactcgtatttattt 1910
               | ||  |||  |||    |   ||    || ||| |  |   |  |  || |
Db    143 CAATATTTAATTATTAAATATAAATAAATCTGTATCTTTATTAGTGTGTTTTTATTCTAA 202

Qy   1911 tcgtacggttgtcgagctcaaaaataaaatcaaattaagctaaaaaaaaaaaaaaaaaaa 1970
            ||   || |  |    ||||| ||||  |||  ||   ||||||||||||||||||
Db    203 AATTAACTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 262


RESULT   9
CNS008EV/c
LOCUS       CNS008EV      946 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR17E23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL052494
VERSION     AL052494.1  GI:4933639
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 946)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
```

```
          NY. The library is named RPCI-98 and was constructed by partial
          EcoRI digestion of Drosophila DNA provided by the BDGP from the
          isogenic strain y2; cn bw sp, the same strain used for the BDGP's
          P1 and EST libraries. A more detailed description of the library
          and how to order individual BAC clones, the entire library, or
          filters for hybridization from the BACPAC Resource Center can be
          found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .946
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR17E23"
                     /note="end : T7"
BASE COUNT      200 a    169 c    226 g    302 t     49 others
ORIGIN


  Query Match             2.3%;  Score 45.8;  DB 122;  Length 946;
  Best Local Similarity   61.1%;  Pred. No. 0.037;
  Matches   69;  Conservative    3;  Mismatches   41;  Indels    0;  Gaps    0;

Qy   1858 caaagattcgatagctttttgtcatagtccttagtcttaactcgtatttatttttcgtacg 1917
           |||   ||| | |  |  |||| |||||    : | ||||    | | |   ||| ||
Db    329 CAACTGTTCAAAAAATCCTTGTTATAGTTAAGWTTTTTAAACTTTTTGTTAGTTCTTATA 270

Qy   1918 gttgtcgagctcaaaaataaaatcaaattaagctaaaaaaaaaaaaaaaaaaa 1970
           |  ||    | |||||||| |||: |||: ||||||||||||||||||||
Db    269 CAAGAAGATTCAATAAATAAAAGCAARGTAAVATAAAAAAAAAAAAAAAAAA 217


RESULT  10
C24663
LOCUS       C24663         455 bp    mRNA            EST       28-APR-1999
DEFINITION  C24663 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SL-X066, mRNA sequence.
ACCESSION   C24663
VERSION     C24663.1  GI:2243084
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
  MEDLINE   99156227
COMMENT     On Jul 9, 1999 this sequence version replaced gi:5435125.
            Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
            Email: d402hu@sakura.cc.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .455
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SL-X066"
                     /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
                     /dev_stage="slug"
BASE COUNT      262 a     56 c     20 g    117 t
ORIGIN


  Query Match             2.3%;  Score 45.4;  DB 80;  Length 455;
  Best Local Similarity   56.3%;  Pred. No. 0.038;
  Matches   85;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy   1820 tcacgaaacacacaaaaacacagcgaattgtagcacctcaaagattcgatagctttttgt 1879
           |||| ||| | | ||||| | |   ||   | | |   |||| |  |  |   |||| |
Db    265 TCACCAAAAAAAAAAAAAAAAAAAAAAAAATGGAAAACCAAAAAACCCTTTTTTTTTTTT 324

Qy   1880 catagtccttagtcttaactcgtatttatttttcgtacggttgtcgagctcaaaaataaaa 1939
             |  |||| |       |   |   | |||   |   ||    | | | |||| ||||
Db    325 TTTTTTCCTTGGAACCCTTTAAAAAAAAATTTAAAATTTTTAAAAACCCCCAAAAAAAAA 384

Qy   1940 tcaaattaagctaaaaaaaaaaaaaaaaaa 1970
             |||  ||   ||||||||||||||||||
Db    385 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 415


RESULT  11
CNS00DKY/c
LOCUS       CNS00DKY      928 bp    DNA             GSS       04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL071865
VERSION     AL071865.1  GI:4948170
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 928)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .928
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR27A24"
                     /note="end : T7"
BASE COUNT      262 a     70 c     84 g    321 t    191 others
ORIGIN


  Query Match             2.3%;  Score 45.2;  DB 122;  Length 928;
  Best Local Similarity   28.9%;  Pred. No. 0.054;
  Matches   41;  Conservative   55;  Mismatches   46;  Indels    0;  Gaps    0;

Qy   1829 acacaaaaacacagcgaattgtagcacctcaaagattcgatagctttttgtcatagtcct 1888
           | : :|:::|::: : ::    : ::: |:::: :::: :::  |||||  :  |  |  |
Db    853 AAMAMAMMMCVMMMMMMMMMMMMMMMTTMWWHHHWHHHHHHWTTTTTTTTWTTTTTTTTT 794

Qy   1889 tagtcttaactcgtatttatttttcgtacggttgtcgagctcaaaaataaaatcaaattaa 1948
           |  | :::: :  ::|||::::|  ::   :: |     |::|:||::::: | |||  :
Db    793 TTTTWWWWWWWWWWWTTTWWWWTWWWWWWWWWWTNNNWNMCHWAWATWWWWWWATATTTW 734
```

```
Qy   1949 gctaaaaaaaaaaaaaaaaaa 1970
          |: :|:| |:||||||||||
Db    733 TATWTWAWATAWAAAAAAAAA 712


RESULT  12
AU071134
LOCUS       AU071134      311 bp    mRNA            EST       10-JUN-1999
DEFINITION  AU071134 Rice cDNA from young root Oryza sativa cDNA clone
            R10796_2Z, mRNA sequence.
ACCESSION   AU071134
VERSION     AU071134.1  GI:5039024
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE   1  (bases 1 to 311)
  AUTHORS   Yamamoto,K. and Sasaki,T.
  TITLE     Rice cDNA from young root
  JOURNAL   Unpublished (1999)
COMMENT     On May 7, 1998 this sequence version replaced gi:3118821.
            Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .311
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="R10796_2Z"
                     /clone_lib="Rice cDNA from young root"
                     /tissue_type="young root"
BASE COUNT      129 a     55 c     44 g     82 t      1 others
ORIGIN


  Query Match             2.3%;  Score 44.6;  DB 48;  Length 311;
  Best Local Similarity   54.6%;  Pred. No. 0.058;
  Matches   89;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy   1808 gctatttcaccttcacgaaacacacaaaaacacagcgaattgtagcacctcaaagattcg 1867
          ||||||    |      | | | | || |  ||| ||| |||  |     || |    ||
Db    127 GCTATTCTTACGCGGGGCACCTCGCAGAGCCACTGCGTATTCAATATTGTCTAGCTATCT 186

Qy   1868 atagctttttgtcatagtccttagtcttaactcgtatttatttttcgtacggttgtcgagc 1927
          |     || | |||   | |    ||| |   | || | |  ||  |  |  | || |
Db    187 ACTCTATTGTATCAGTATTCAGTATCTCATTGCTTAATAACATTATTCAGCATCTCAAAG 246

Qy   1928 tcaaaaataaaatcaaattaagctaaaaaaaaaaaaaaaaaa 1970
          |||||| ||||  |||  ||   ||||||||||||||||||
Db    247 ACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 289


RESULT  13
CNS00BTL/c
LOCUS       CNS00BTL      825 bp    DNA             GSS       04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR24E04 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL058023
VERSION     AL058023.1  GI:4938085
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 825)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .825
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR24E04"
                     /note="end : TET3"
BASE COUNT      178 a    132 c    188 g    297 t     30 others
ORIGIN


  Query Match             2.3%;  Score 44.6;  DB 122;  Length 825;
  Best Local Similarity   60.2%;  Pred. No. 0.078;
  Matches   68;  Conservative    3;  Mismatches   42;  Indels    0;  Gaps    0;

Qy   1858 caaagattcgatagctttttgtcatagtccttagtcttaactcgtatttatttttcgtacg 1917
          |||   ||| | |  |  ||:| |||||    | | ||||     |:|   ||: ||
Db    312 CAACTKTTCAAAAAATACTTKTTATAGTTAAGATTTTTAAACTTATTKTTAGTTYTTATA 253

Qy   1918 gttgtcgagctcaaaaataaaatcaaattaagctaaaaaaaaaaaaaaaaaaa 1970
          |  ||    | |||||||| |||| |||  ||||||||||||||||||
Db    252 MAAGAAGATTCAATAAATAAAAGCAAAGTAAAATAAAAAAAAAAAAAAAAAA 200


RESULT  14
AA839308
LOCUS       AA839308      315 bp    mRNA            EST       27-FEB-1998
DEFINITION  vo40d03.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA
            clone IMAGE:1052357 5', mRNA sequence.
ACCESSION   AA839308
VERSION     AA839308.1  GI:2915403
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 315)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Feb 17, 1998 this sequence version replaced gi:2889382.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
```

```
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:583933
            Putative full length read
            vector to vector length is 316
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .315
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1052357"
                     /clone_lib="Barstead mouse irradiated colon MPLRB7"
                     /dev_stage="8 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained
                     from 8 week old mouse.  Colon was harvested 72 hours after
                     irradiation with 1400 Gys.  1st strand cDNA was primed
                     with  a Not I - oligo(dT) primer
                     [5'TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors [AATTCGGATCCTTG], digested with Not I and cloned
                     into the Not I and Eco RI sites of  the modified pT7T3
                     vector. Library constructed by Bob Barstead.  "
BASE COUNT      108 a     69 c     43 g     95 t
ORIGIN


  Query Match             2.3%;  Score 44.4;  DB 31;  Length 315;
  Best Local Similarity   59.5%;  Pred. No. 0.067;
  Matches   75;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy   1845 aattgtagcacctcaaagattcgatagctttttgtcatagtccttagtcttaactcgtat 1904
           | ||||||||||||| | || || | |   ||   || |   ||  ||| ||| |    | |
Db    185 ATTTGTAGCACCTCTAGGACTCCAAAATTTGGAGTTAACCTCTCTAGCCTTCAGAATTGT 244

Qy   1905 ttattttcgtacggttgtcgagctcaaaaataaaatcaaattaagctaaaaaaaaaaaaa 1964
           |||| |   |  |  |  | |  |  |||| |  || | |       ||||||||||||
Db    245 TTATATATCTGTGTATAGCTATATATAAAACATCATTATAAACTAAAAAAAAAAAAAAA 304

Qy   1965 aaaaaa 1970
           ||||||
Db    305 AAAAAA 310


RESULT  15
AW307211
LOCUS       AW307211      202 bp    mRNA            EST       20-JAN-2000
DEFINITION  sf54b11.y1 Gm-c1009 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1009-3766 5', mRNA sequence.
ACCESSION   AW307211
VERSION     AW307211.1  GI:6719564
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales;
            Fabaceae; Papilionoideae; Glycine.
REFERENCE   1  (bases 1 to 202)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .202
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1009-3766"
                     /clone_lib="Gm-c1009"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; The mRNA was isolated from entire roots of
                     2-month-old 'Williams' plants that were greenhouse grown
                     in 5-gallon pots. To suppress nodulation, Black Gold
                     All-Purpose potting soil was supplemented with: 0.36g/L
                     available phosphoric acid (P2O5), 20mg/L urea N, 0.16g/L
                     S, 0.49mg/L B, 2.5mg/L Cu, 0.15g/L Fe, 13.53mg/L Mn,
                     0.26mg/L Mo, 14mg/L Zn, 20mg/L Ca, and the following
                     nutrients in a slow-release form (Osmocote): 0.165g/L
                     ammonia N, 0.185g/L nitrate N, 0.35g/L available
                     phosphoric acid, and 0.35g/L soluble potash. No nodules
                     were visible on the roots at harvest. Stratagene's cDNA
                     Synthesis Kit (catalog #200401) was used to synthesize the
                     cDNA. First-strand synthesis was performed with 5-methyl
                     dCTP, hence the ligated cDNA is hemimethylated.
                     Stratagene's first-strand synthesis primer was used
                     [GAGAGAGAGAGAGAGAGAACTAGTCTCGAG(T)-18]. After
                     second-strand synthes is, the cDNA ends were 'polished'
                     with clone Pfu DNA polymerase, ligated to EcoRI adapters,
                     and phosphorylated. The XhoI site within the first-strand
                     synthesis primer was restricted by digestion with XhoI;
                     all XhoI sites in the cDNA would be protected by their
                     hemimethylated status. The cDNA constructs were
                     size-fractionated with a 400bp cutoff, using a SizeSep 400
                     Spun column from Pharmacia. The column eluent was then
                     ligated into Stratagene's pBluescript II XR Predigested
                     vector (pBluescript II SK(+) that had been digested with
                     EcoRI and XhoI, and phosphorylated). Both the white and
                     blue colonies appear to contain recombinant plasmids with
                     cDNA inserts. This library was constructed by Dr. Paul
                     Keim and Dr. Virginia Coryell."
BASE COUNT      102 a     22 c     22 g     56 t
ORIGIN


  Query Match             2.2%;  Score 43.8;  DB 70;  Length 202;
  Best Local Similarity   64.1%;  Pred. No. 0.087;
  Matches   66;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy   1868 atagctttttgtcatagtccttagtcttaactcgtatttatttttcgtacggttgtcgagc 1927
           ||  | |||||  ||| |  ||   | ||||   || ||   | |||    | || ||
Db     73 ATGTCGTTTTGAAATATTATTTGAACCTAACAATTAATTTCATGCGTTAATTCGTTGAAG 132

Qy   1928 tcaaaaataaaatcaaattaagctaaaaaaaaaaaaaaaaaaa 1970
           | |||||  ||||  |||  ||   ||||||||||||||||||
Db    133 TAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 175



Search completed: September  4, 2000, 00:24:48
Job time: 19863 sec
```

THIS PAGE BLANK (USPTO)